

Univerzita Karlova v Praze

Pedagogická fakulta

# **Automatické opravy v prostředí MS Office Word a jejich vliv na gramatiku mládeže**

Jaroslav Král

Katedra informačních technologií a technické výchovy

Vedoucí bakalářské práce: PhDr. Jiří Štípek Ph.D.

Studijní program: B7507 Specializace v pedagogice

2012



UNIVERZITA KARLOVA V PRAZE

PEDAGOGICKÁ FAKULTA

Katedra informačních technologií a technické výchovy

## ZADÁNÍ BAKALÁŘSKÉ PRÁCE

akademický rok 2010/2011

Jméno a příjmení studenta: **Jaroslav Král**

Studijní program: **B7507 Specializace v pedagogice**

Studijní obor: **Informační technologie se zaměřením na vzdělávání**

Název tématu práce v českém jazyce:

**Automatické opravy v prostředí MS Office Word a jejich vliv na gramatiku mládeže.**

Název tématu práce v anglickém jazyce:

**Automatic corrections in MS Office Word and their impact on children's grammar.**

Pokyny pro vypracování:

- Charakterizovat systémy oprav textu soudobých textových procesorů.
- Analyzovat a zhodnotit vybraný systém kontroly pravopisu a gramatiky.
- Na základě studia příslušných odborných pramenů popsat aspekty využití uvedených systémů a jejich dopad na gramatiku školní mládeže.
- Experimentálně zjistit míru závislosti gramatické úrovně textů školní mládeže na využití systémů kontroly pravopisu a gramatiky.
- Zhodnotit schopnosti školní mládeže obejít se bez uvedených systémů a případně navrhnout a zdůvodnit orientaci dalších výzkumných aktivit

Vedoucí bakalářské práce: PhDr. Jiří Štípek Ph.D.

Předpokládaný rozsah bakalářské práce<sup>1</sup>: 40 normostran

Datum zadání práce: 12. 09. 2011

Předběžný termín odevzdání práce: 15. 06. 2012

Práce se odevzdává ve dvou knihařsky svázaných exemplářích v pevných deskách. Současně se odevzdává jeden její stejnopis na nepřepisovatelném nosiči dat (CD, DVD).

V Praze dne:

.....  
doc. PhDr. Vladimír Rambousek, CSc.  
vedoucí katedry

---

<sup>1</sup> Minimální rozsah bakalářské práce činí standardně 40 normostran (72 000 znaků vč. mezer) vlastního textu.

Prohlašuji, že jsem bakalářskou práci na téma Automatické opravy v prostředí MS Office Word a jejich vliv na gramatiku mládeže vypracoval pod vedením vedoucího bakalářské práce samostatně za použití v práci uvedených pramenů a literatury. Dále prohlašuji, že tato bakalářská práce nebyla využita k získání jiného nebo stejného titulu.

15.06.2012

.....

podpis

Rád bych touto cestou vyjádřil poděkování PhDr. Jiřímu Štípkovi Ph.D. za jeho cenné rady a trpělivost při vedení mé bakalářské práce. Rovněž bych chtěl poděkovat Mgr. Ivanu Vargovi za vstřícnost a pomoc při získání potřebných informací a podkladů. Poděkování dále také patří vedení a žákům Plaského gymnázia za umožnění provedení praktické části bakalářské práce.

.....

podpis

**NÁZEV:**

Automatické opravy v prostředí MS Office Word a jejich vliv na gramatiku mládeže

**AUTOR:**

Jaroslav Král

**KATEDRA (ÚSTAV):**

Katedra informačních technologií a technické výchovy

**VEDOUCÍ PRÁCE:**

PhDr. Jiří Štípek Ph.D.

**ABSTRAKT:**

Ve své teoretické části se tato bakalářská práce zabývá představením základních principů fungování prostředků kontroly pravopisu a gramatiky v textových procesorech. Dále je krátce zmíněna historie textových procesorů a také základy lingvistiky nutné k pochopení fungování principů oprav textu. V praktické části bakalářské práce je prováděn vlastní výzkum. Jeho cílem je zjistit, zda existuje vazba mezi využitím prostředků automatických oprav a oprav pravopisu a gramatiky v prostředí textového procesoru Microsoft Office Word 2007 a úrovni gramatického projevu mládeže. Dalším cílem je zhodnocení rozložení chyb v rukou psaném textu z hlediska jejich rozdělení na chyby rozpoznávané procesorem MS Office Word 2007 a chyby tímto procesorem nerozpoznávané. Pro získání dat nutných k vyhodnocení výzkumu je použito dotazníku a dále diktátu, který je psán ve dvou shodných verzích nejprve v ruce a poté pomocí Microsoft Office Word 2007. Vyhodnocení výzkumu přináší některé zajímavé výsledky, nicméně hypotézy výzkumu nejsou potvrzeny.

**KLÍČOVÁ SLOVA:**

Textové procesory

Microsoft Office Word 2007

Gramatika mládeže

**TITLE:**

Automatic corrections in MS Office Word and their impact on children's grammar

**AUTHOR:**

Jaroslav Král

**DEPARTMENT:**

Information Technology and Technical Education Department

**SUPERVISOR:**

PhDr. Jiří Štípek Ph.D.

**ABSTRAKT:**

In its theoretical part this bachelor thesis focuses on introducing how spelling and grammar corrections in text processors work on basic level. Brief history of text processors as well as linguistic basics needed for comprehension of the discussed problems are mentioned. In its practical part the bachelor thesis focuses on a research. Its goal is to examine whether there is a connection between using grammar and spelling correction in Microsoft Office Word 2007 and the level of grammatical expression of the youth. Another goal is to examine the hand written text conducted in the research for distribution of errors recognized by Microsoft Office Word 2007 in relation to errors not recognized by this program. A questionnaire and a dictate, which is first written in hand and then later in same form using Microsoft Office Word 2007, are used to gather the data necessary for the research. The evaluation of the research reveals some interesting results, however the hypotheses of the research are not validated.

**KEYWORDS:**

Text processors

Microsoft Office Word 2007

Youth's grammar

# Obsah

1	Úvod .....	8
2	Teoretická část .....	9
2.1	Historie .....	9
2.2	Textové editory vs. textové procesory .....	10
2.3	Lingvistika .....	11
2.3.1	Typologie jazyka podle pravidel tvoření slov .....	11
2.3.2	Důležité lingvistické disciplíny .....	12
2.4	Systémy kontroly textu a jeho automatických oprav v textových procesorech .....	15
2.4.1	Identifikace chyby.....	15
2.4.2	Nabízení řešení .....	20
2.5	Představení textových procesorů z hlediska systémů oprav textu .....	23
2.5.1	OpenOffice.org Writer.....	23
2.5.2	Google Documents (docs.google.com) .....	25
3	Praktická část.....	26
3.1	Vlastní výzkum.....	26
3.1.1	Vymezení cíle a hypotéz výzkumu.....	26
3.1.2	Zdůvodnění výběru vzorku a jeho charakteristika a popis.....	27
3.1.3	Plán výzkumu.....	28
3.1.4	Postup při realizaci výzkumu .....	32
3.1.5	Kontrola odpovědí.....	32
3.1.6	Výsledky.....	34
3.2	Interpretace výsledků výzkumu a jeho zhodnocení.....	45
4	Závěr .....	49
5	Seznam použitých informačních zdrojů .....	50
6	Přílohy.....	52

# 1 Úvod

Tématem této bakalářské práce jsou systémy automatických oprav textu a oprav pravopisu a gramatiky v textových procesorech a vliv jejich využívání na psanou komunikaci. Práce se pokusí o stručné rozdělení těchto systémů a vysvětlení principu jejich fungování. V praktické části bakalářské práce je prováděn vlastní výzkum, který zkoumá, zda existuje spojitost mezi formou psaného projevu a úrovní gramatického projevu studentů. Je možné, že prostředí textových procesorů a jejich automatické opravy a opravy pravopisu a gramatiky zhoršují gramatiku, nebo naopak žáky na chyby upozorňují a tím přispívají k jejich zapamatování a v důsledku ke zlepšení gramatické úrovně psaného projevu studentů?

Nacházíme se v době, kdy elektronická komunikace je již naprostou samozřejmostí. Poslední generace, která nastoupila do škol po roce 2004, se již narodila do prostředí s masivním využitím IT technologií a tím je jimi ovlivňována od samého začátku. Jde tak o vzorek populace, který jako první nemůže porovnat dobu před rozšířením IT pro širokou veřejnost s nynější dobou, kdy jsou prostředky IT využívány kýmkoli. Je vystaven vlivům tohoto prostředí v maximální míře a považuje to za naprostou přirozenost. Proto je nyní ideální čas k výzkumům v této oblasti, máme možnost pozorovat něco nového a do jisté míry naprosto revolučního.

Navíc některé prostředky IT mají dnes určité stigma, často je uváděn jejich možný negativní vliv na mládež, mají ji údajně odvádět od aktivního a plnohodnotného života. Jedním z názorů je, že přílišná závislost na informačních technologiích činí děti lenivější a učí je, že nemusí samy přemýšlet, informace si přece naleznou na počítači či internetu. Tento názor může být předpojatý, a proto si tato práce dává za cíl zkoumat jeden z možných dopadů prostředků IT na mládež, a to konkrétně v oblasti psané komunikace.



## 2 Teoretická část

Teoretická část vychází převážně pouze z anglické literatury. Ta se samozřejmě zabývá problémem zpracování a vyhodnocení textu a jeho následných oprav z pohledu angličtiny a ne češtiny, což by se mohlo zdát nevýhodné. Nicméně anglická literatura na toto téma je mnohem dostupnější a tak dobře poslouží pro ilustraci a vysvětlení základních obecných principů automatických oprav. Je nutné si ale také zároveň uvědomit, že každý jazyk má svá specifika. Proto se systémy automatických oprav pro každý jazyk mohou v detailech odlišovat, základní logika fungování jazyka je ale totožná (pokud vynecháme extrémy v podobě kmenových jazyků a dalších podobných systémů verbální komunikace) a tudíž i systémy jeho oprav jsou srovnatelné.

### 2.1 Historie

V dnešní době je prakticky nemožné narazit na textový procesor, který by neumožňoval určitou formu identifikace chyb v textu. Systémy automatických oprav se staly samozřejmou součástí těchto programů. Hlídnou za nás naše překlady a pomáhají nám napravovat chyby způsobené naší nepozorností a spěchem. Jak jsme se ale k tomuto stavu dostali?

První systémy pro pomoc při opravě textu vznikly jako součást procesu vzniku jiných programů či produktů, ne jako jejich přímý účel a výsledek samostatného výzkumu a vývoje. Motivací pro jejich vznik byla potřeba kontroly dat na vstupu programů. Prvotní úsilí bylo směřováno na nalezení a opravu chyb vzniklých na specifických vstupních zařízeních v určitém kontextu. Šlo například o nalezení špatně zapsaných jmen pasažérů na určitých letech nebo přímo o kontrolu zadání jmen v rozhraní odbavovacího systému. V jiném případě se vývojáři zabývali kontrolou jmen míst a lokalit v genealogické databázi. Ve své prvotní fázi byly tedy systémy oprav textu brány jako součást nadřazených programů a ne jako samostatné softwarové nástroje (Peterson, 1980, s. 677).

Podle Lesa Earnesta lze za zcela první program zabývající se kontrolou pravopisu považovat podproces vyvinutý v roce 1961 jako součást aplikace na rozpoznání rukou psaného textu. Ten používal slovník čítající 10 000 anglických slov. Text byl se slovníkem porovnáván a podprogram poté nabízel podle shody slova, která rukou psaný text reprezentuje (Earnest, 2011, s. 1). Proces tedy sloužil jako oprava chyb napsaných v rukou psaném textu nebo ještě s větší pravděpodobností pro opravu chyb vzniklých při analýze převáděného textu.

Les Earnest se této problematice věnoval dále a jako člen Stanford Artificial Intelligence Lab (Stanfordská laboratoř umělé inteligence) pověřil v roce 1967 studenta vytvořením programu

pro kontrolu textových souborů.<sup>2</sup> Ten opět využíval Earnestův slovník a v textu vyhledával slova, která v něm nebyla uvedena. Program vygeneroval textový soubor obsahující seznam všech nerozpoznaných slov spolu s číslem stránky a řádku, na kterém se vyskytují (Earnest, 2011, s. 1).

Prvním samostatně vyvíjeným programem na kontrolu textu, který nebyl součástí jiné aplikace, je program SPELL napsaný studentem Stanford University Ralphem Gorinem v roce 1971 (Peterson, 1980, s. 677). Vznikl na žádost opět Lesa Earnesta. SPELL stále porovnával slova s Earnestovo slovníkem, nicméně tentokrát umožňoval i interaktivní opravy. Pokud nenašel ve slovníku shodu, pokusil se totiž najít alespoň slova, která se liší pouze v jednom písmenu a poté je nabídl jako možnou opravu (Earnest, 2011, s. 1).

V dnešní době existují systémy na kontrolu pravopisu spíše jako součást textových procesorů než jako samostatné programy. Vše se změnilo s příchodem osobních počítačů s dostatečným výkonem v začátku osmdesátých let dvacátého století. Mezi používané textové procesory patřil například WordPerfect či v té době jeho přímý konkurent a nynější gigant Microsoft Word (Eisenberg, 1992, s. 268-278).

## 2.2 Textové editory vs. textové procesory

Než budou rozebrána jednotlivá prostředí, je třeba upřesnit dva termíny, resp. slovní spojení, která jsou v praxi chápána poměrně volně a někdy jsou dokonce zaměňována. Jedná se o „textové procesory“ a „textové editory“. Může se zdát, že jde o naprosto zástupné označení stejného typu programů, to je ale mylný pohled. Mezi oběma je poměrně zásadní rozdíl.

**Textové editory** slouží pro pořizování a úpravu textů (např. strukturace textu, tvorba nadpisů, volba typů písem apod.) (Balvínová, [cit. 2012-03-14a]). Zaměřují se tedy převážně na prostý text a vytvoření obsahu. Mezi nejznámější zástupce textových editorů patří například poznámkový blok v Microsoft Windows. Za textový editor lze také považovat programy PSPad, Notepad 2, Notepad ++ a mnohé další aplikace orientované především na text, s limitovanými možnostmi jeho formátování.

**Textové procesory** naproti tomu umožňují také složitější úpravy, formátování a tisk dokumentů. Rozšiřují možnosti textových editorů o další funkce, např. o umísťování obrázků, tabulek nebo jiných objektů do textu, hromadné vyhledávání či nahrazování, vytváření záhlaví a zápatí, automatické dělení slov, kontrolu pravopisu. Lze jimi vytvářet různé šablony,

---

<sup>2</sup> Zajímavým faktem je, že student tento program napsal údajně během dvou dnů (Earnest, 2011, s. 1).

volit formát stránky, exportovat v různých datových formátech, popř. vytisknout na různých tiskárnách (Balvínová, [cit. 2012-03-14b]).

Důležitou odlišností textových procesorů od editorů je také jejich takzvaný WYSIWYG charakter. Jedná se o zkratku z anglické věty „What you see is what you get“ znamenající přeneseně „Co vidíš, to dostaneš“. Jde o to, že v procesorech je text rovnou zobrazen v jeho formátované podobě, tedy měli bychom vidět přímo to, co bude vytištěno v tiskárně (Smith, 2009, s. 152).

## 2.3 Lingvistika

Pokud má systém automatických oprav textu v rámci svých možností správně fungovat, musí z logiky věci alespoň do určité míry respektovat a rozeznávat strukturu a zákonitosti fungování jazyka. Již první program na kontrolu textových souborů vyvinutý v roce 1967 (viz 2.1 Historie) respektoval pravidla anglické morfologie (tvarosloví neboli morfologie je nauka o slovních druzích, o jejich tvarech, jakož i o významech tvarů slov.<sup>3</sup> Program byl schopen pomocí speciálního algoritmu oddělit od slova přípony a předpony a dále pracoval již jen s jeho kořenem (Earnest, 2011, s. 1). To umožnilo programu se svou slovníkovou databází porovnávat i slova odvozená a tím se zvýšila jeho efektivita. Původně se předpokládalo, že takový přístup může mít za následek přehlédnutí chyb, které se v příponách a předponách vyskytují, avšak tato obava se v praxi nepotvrdila (Earnest, 2011, s. 1).

V dalších podkapitolách budou nastíněny některé ze základních relevantních teoretických zákonitostí jazyka. Tyto zákonitosti, resp. pravidla, jsou základem principů fungování systémů na opravu textu a logika jejich fungování vychází z poznatků lingvistiky.

### 2.3.1 Typologie jazyka podle pravidel tvoření slov

Jedním z možných přístupů jak kategorizovat jazyky je jejich dělení na základě jejich pravidel pro vytváření slov. Následující text pak bude na toto dělení několikrát odkazovat. Dělení je převzato z knihy o teorii jazyka *How Language Works* od Davida Crystala.

- Izolační jazyky (také nazývané analytické jazyky)
  - Všechna slova v těchto jazycích jsou neměnná: nejsou zde žádné koncovky, skloňování. Gramatické vztahy mezi slovy jsou tak realizovány pouze pomocí jejich pozice ve větě. Jako příklad můžeme použít Čínštinu či Větnamštinu.

---

<sup>3</sup> JIHOČESKÁ UNIVERZITA V ČESKÝCH BUDĚJOVICÍCH, PedF, KATEDRA ČESKÉHO JAZYKA A LITERATURY. *ZÁKLADY ČESKÉ MORFOLOGIE: Základní teze, terminologie a kategorie tématu*

- Inflektivní jazyky (také nazývané syntetické jazyky)
  - Gramatické vztahy jsou vyjádřeny pomocí změny interní struktury slov – často za použití skloňování (předpony, přípony, vpony). Latina, Řečtina a Arabština jsou zástupci této skupiny. Mezi invektivní jazyk patří také Čeština.
- Aglutinační jazyky
  - Budují slova pomocí dlouhých sekvencí jednotek kde každá z nich má určitý gramatický význam. Turečtina, Japonština a Finština tvoří slova touto cestou.
- Polysyntetické jazyky
  - Tvoří většinou dlouhá a komplexní slova obsahující inflektivní i aglutinační prvky. Jedná se často o kmenové jazyky, jako Eskimo či jazyk Mohawk .

Jak je vidět, v některých případech jsou rozdíly celkem značné. Slova mohou být velice jednoduchá, jako například anglické „a“, nebo naopak velice komplexní jako „palyamunurringkutjamu nurtu“ znamenající v jednom z australských domorodých jazyků „ten/ta se rozhodně nestal/la zlým/zlou“ (Crystal, 2007, s. 240). Z hlediska programů oprav textu není nejspíš nutné brát v potaz jazyky afrických křováků či Eskymáků, je ale vhodné logiku fungování programu uzpůsobit zákonitostem jazyka. Pro každý konkrétní jazyk je tak příhodné systém automatických oprav upravit.

### 2.3.2 Důležité lingvistické disciplíny

Z hlediska prvnopisu, gramatiky a pozice slova ve větě jsou zásadní zejména tyto lingvistické disciplíny:

- morfologie
  - Tvarosloví (morfologie) je část gramatiky, která se zabývá tvořením tvarů slov. Pojednává o slovních druzích, jejich tříděních, o tvarech slov a jejich mluvnických významech (kategoriích) (Dobešová, Fialová, 2005, s. 27).
  - Jak již bylo ukázáno výše na příkladu programu z roku 1967, respektování pravidel morfologie rozšiřuje využitelnost porovnávání se slovníkem pojmů. Nicméně záběr morfologie nepřesahuje hranice jednoho konkrétního slova, nezabývá se vztahy mezi více slovy či dokonce celými větami (Crystal, 2007, s. 236).

- Morfologie jazyka je důležitá pro programy rozpoznávání textu také z důvodu identifikace jednotlivých slov, s kterými může poté program pracovat jako s jednotkami významu. Je tak velice kritická zejména pro aglutinační a inflektivní jazyky, kde diktuje jejich správnou gramatickou formu slov.
- syntax
  - Skladba (syntax) je jazykovědná disciplína, která se zabývá tvořením vět a souvětí, jejich strukturou a využíváním v komunikaci (Dobešová, Fialová, 2005, s. 57). Syntaxi můžeme chápat jako systém nadřazený morfologii, zabývá se strukturou vět a souvětí, ne jednotlivých slov, jako morfologie (Crystal, 2007, s. 247).
  - Jak již bylo řečeno výše, syntaxe se zabývá větami. Implementace jejich pravidel do programů kontroly textu rozšiřuje jejich záběr. Nejsou pak schopny pouze řešit problémy na úrovni jednotlivých slov (překlepy, pravopisné chyby), ale také například slovosled, psaní interpunkce a tak dále. Jak si můžeme všimnout například při používání Microsoft Word 2007, textový procesor je již schopen nabízet opravy interpunkce při psaní souvětí.
- lexikologie
  - Lexikologie může být definována jako věda studující slovní zásobu jazyka. Jako taková si dává za cíl také vytvoření takzvaného lexikonu jazyka, což je souhrn jeho celkové slovní zásoby (Jackson, Zé Amvela, 2000, s. 1). Lexikologie tak může přispět při vytváření databáze slov uložených ve slovnících využívaných programy kontroly textu.
- sémantika
  - Sémantika je obecně definována jako studie významu. Dále se dělí na pragmatickou sémantiku, studující význam vět v kontextu prostředí, kde byly použity; větnou sémantiku zabývající se významem vět a také vztahy mezi nimi na základě významu a na lexikální sémantiku zkoumající významy slov (Jackson, Zé Amvela, 2000, s. 4).
  - Sémantika tedy zkoumá věty a slova z hlediska jejich významu. Zamezí nám například v použití věty „Klávesnice snědla létající telefonní budku.“ Tato věta je z hlediska morfologie v pořádku, syntaxe je také podle pravidel (slovosled a

povinné větně členy jsou dodrženy). Nicméně z hlediska sémantiky je tato věta naprosto bez významu a tím pádem postrádá její použití v psaném textu nebo při verbální komunikaci smysl.

V ideálním případě by měly systémy automatických oprav textu respektovat a rozeznávat všechny tyto lingvistické disciplíny a jejich pravidla, a ve skutečnosti mnohé další. Poté by se staly plnohodnotným softwarovým vybavením nevyžadujícím následnou kontrolu či asistenci uživatele v procesu oprav. Již nyní se můžeme setkat s prvními pokusy, nicméně jejich funkčnost je zatím omezená, jak může ilustrovat následující ukázka z básně „Candidate for a Pullet Surprise“ od Jerrolda H. Zara. Ta byla vytvořena speciálně pro ilustraci nedokonalosti automatických oprav anglického jazyka (Nordquist, [cit. 2012-03-24]).

*„To rite with care is quite a **feet**  
Of **witch** won should **bee** proud,  
And we **mussed** dew the best **wee** can,  
Sew flaw's are **knot** aloud.“*

Báseň využívá principu homonymie (homonymie = stejné znění slov, přípon nebo koncovek různého významu<sup>4</sup>). Navíc postrádá pravopisné chyby a překlepy, veškerá slova ve své individuální podobě v jazyce existují. Ukázka dává smysl, ale až poté, co se odpoutáme od psané podoby slov a zaměříme se pouze na zvukovou složku. Poté získáme:

*„Psát s citem je celkem výkon  
na který bychom měli být hrdí,  
a musíme dělat co je v našich silách,  
aby chyby nebyly umožněny.“*

Doslovně, tak jak je napsaná, ale znamená (spojení „to rite“ v překladu ignoruji, jelikož jednoduše nelze přeložit, nedává v logice angličtiny smysl, nemůže existovat. „To“ uvozuje infinitiv slovesa, „rite“ je ale podstatné jméno; spojení slov „to“ a „rite“ by bylo možné – musel by ale mezi nimi ještě být určitý či neurčitý člen):

*„...s citem je celkem chodidla  
které čarodějnice vyhrála měla by včela hrdá,  
a my rozcuchali rosu nejlépe malý moci,  
sešít chyby jsou uzel nahlas.“*

---

<sup>4</sup> Homonymie. In: [Http://slovník-cizich-slov.abz.cz](http://slovník-cizich-slov.abz.cz) [online]. c 2005-2006 [cit. 2012-04-05]. Dostupné z: <http://slovník-cizich-slov.abz.cz/web.php/slovo/homonymie>

Pokud vycházíme z předpokladu, že báseň byla při svém vzniku v roce 1992 (Nordquist, [cit. 2012-03-24]) napsána tak, aby žádná z chyb nebyla identifikována, je pak vidět pokrok v systému identifikace chyb. Tato bakalářská práce je psána v programu Microsoft Word 2007 a ten již v anglickém znění identifikoval šest chyb a nabídl správné opravy. Přitom slova jsou napsána bez pravopisných chyb, chybné je pouze jejich užití ve větě. Slova, u kterých byla nabídnuta jiná forma, jsou zvýrazněna žlutým pozadím. I tak ale v anglické ukázce zůstává šest chyb. Lze tedy přepokládat, že implementace pravidel syntaxe a sémantiky do oprav anglického textu se zlepšila.

Pro zajímavost, na mém hrubém českém překladu programu vadila pouze chybějící čárka mezi slovy „vyhrála“ a „měla“, jinak byl se zněním ukázky naprosto spokojen.

## **2.4 Systémy kontroly textu a jeho automatických oprav v textových procesorech**

Další část bakalářské práce je zaměřena, jak napovídá nadpis, na samotný princip, nebo spíše jednotlivé principy, fungování systémů oprav. Ve svém popisu se pokusím následovat cestu imaginárního programu při kontrole vstupních dat a postupně ilustrovat překážky, na které může narazit, a činnosti, které mu umožní tyto překonat.

### **2.4.1 Identifikace chyby**

#### ***2.4.1.1 Metody nevyužívající slovník (pracující pod úrovní slova)***

Poměrně známou metodou pro rozeznání chybně napsaného slova, kterou je schopen samostatně identifikovat snad každý pokročilejší uživatel textových procesorů, je takzvaná slovníková metoda. Zmiňovala se o ní již část o historii systému oprav. Nicméně existují i metody, které operují ve struktuře jazyka ještě na nižší úrovni, než jsou slova. Tyto metody slovníky vůbec nepoužívají, nebo pouze nepřímo (Mitton, 1996, s. 93).

Obě při své funkci používají n-gramy. Jde o skupiny písmen, které se nacházejí v sousední pozici v rámci jednoho slova. Většinou jde o samostatné písmeno, dvojici, nebo trojici písmen. Neboli monogramy (či unigramy), digramy (či bigramy) a trigramy. S těmito stavebními kameny slov poté programy dále pracují. Nejdříve je v každém slově identifikují a poté zkoumají pravděpodobnost jejich výskytu, či jestli vůbec jako takové mohou existovat (Kukich, 1992, s. 380).

První metoda používá slovník nepřímo. Celý ho analyzuje a zaznamená všechny v něm se vyskytující trigramy. Následně pracuje s opravovaným textem, rozdělí ho na trigramy a ty porovná se seznamem získaným v předešlém kroku. Pokud nenajde shodu (trigram se ve

slovníku nevyskytoval), jedná se nejspíše o pravopisnou chybu či překlep. Nevýhodou tohoto přístupu je fakt, že pravopisné chyby nemají za následek vždy neexistující trigramy. Proto může tato metoda některé chyby přehlédnout. Je nicméně poměrně efektivní při opravách chyb vzniklých při čtení textu OCR („optical character recogniton“ – optické rozeznávání znaků) programy (Mitton, 1996, s. 93).

Druhá metoda nepoužívá slovník vůbec. Při kontrole dokumentu využívá jako zdroj pro vytvoření referenčního seznamu trigramů samotný dokument. Tento seznam je založený na frekvenci výskytu trigramů. Poté je vypočítán „index zvláštnosti / podivnosti“ (orig. index of peculiarity) pro každé slovo v dokumentu na základě trigramů, které obsahuje. Následně jsou všechna slova sestupně seřazena podle jejich „zvláštnosti“. Očekává se, že špatně napsaná slova budou v horní části tohoto seznamu (Kukich, 1992, s. 380). Výhodou tohoto přístupu je také to, že není omezen jazykem, kterým je napsán zdrojový dokument. U předchozí metody bylo nutné dodat slovník odpovídající jazyku textu, tato ale pracuje čistě na statistické bázi a využívá pouze svůj vstupní soubor.

#### ***2.4.1.2 Metody přímo využívající slovník (pracující na úrovni slova)***

Nejjednodušším přístupem je využívat slovník přímo, aniž by se porovnávaná slova jakkoli upravovala. Program pak funguje na podobném principu jako sám člověk, jednoduše vezme slovo ze zdrojového dokumentu a hledá ho ve slovníku. Ten může být abecedně seřazený, aby se uspíšilo hledání. Pokud slovo ve slovníku není, jedná se o chybnou formu. Nicméně tato metoda má vyšší požadavky na úložný prostor (pro větší efektivnost musí být slova ve slovníku uchována totiž zvláště i ve své skloňované formě) a i její rychlost vyhledávání není ideální, probíhá sekvenčně (Mitton, 1996, s. 94).

Nabízí se tedy možnost využití hashovací tabulky. Slovu zdrojového dokumentu je pomocí hashovací funkce přiřazena adresa v hashovací tabulce. Pokud není slovo uloženo na adrese totožné, nebo adresa obsahuje prázdnou hodnotu, jedná se nejspíše o chybu. Výhodou této metody je její rychlost vyhledávání slova oproti sekvenčnímu přístupu. Nevýhodou je nutnost vytvoření takové hashovací funkce, která by pokud možno eliminovala možnost přiřazení stejného klíče dvěma a více slovům při zachování co nejmenší hashovací tabulky. Oproti předchozí metodě je „hashovací“ rychlejší, její nároky na paměť jsou ale stále relativně vysoké. Nevýhodou je, že i přesto že hashovací adresa je závislá na obsahu a formě slova, je stále možné, aby dvě různá slova vedla ke stejné adrese. Řešením je neuchovávat na adrese v tabulce celé slovo, ale pouze indikovat zda vůbec vede ke korektně napsanému slovu. Na adrese je tak uchováván pouze bit signalizující že vede k validnímu slovu. Pokud ale nyní



nezkoumáme, co je vlastně na adrese uloženo, může být i špatně napsané slovo programem vyhodnoceno jako korektní (Kukich, 1992, s. 382).

Omezená paměť také často motivovala rané programy k vyhnutí se uchovávání všech možných forem slova ve slovníku (skloňování, časování atd.). Místo toho je ve slovníku pouze kořen slova a pokud není zkoumané slovo ve slovníku nalezeno, program se pomocí morfologických pravidel jazyka pokusí odebrat možné přípony a předpony, či je nahradit za jiné a poté znovu hledat shodu se slovníkem. Nicméně ani tento způsob není ideální. Může ignorovat chybu napsanou v předponě či příponě. Také by mohl označit jako správná slova ta, která vznikla připojením špatné přípony či předpony ke kořenu slova. Za příklad může sloužit slovo „*undoubt*“ (Mitton, 1996, s. 94). Jde o spojení negativní předpony „*un*“ a slova „*doubt*“, znamenající „*pochyba*“. Nicméně nově vzniklé slovo není platné, jelikož nedává v anglickém jazyce smysl (slovo by znamenalo „*odpochybyt*“).

#### **2.4.1.3 Metody schopné identifikovat chyby nad rámcem slova**

Předchozí kapitoly se zabývaly výhradně chybami uvnitř jednotlivých slov bez ohledu na jejich lingvistické okolí. Metody identifikace těchto chyb přistupovaly ke slovům jako k samostatným jednotkám bez ohledu na další slova v jejich okolí, či bez ohledu na větu jako celek. Šlo tedy převážně o překlady na klávesnici, pravopisné chyby a tak dále. Jak ale přistupovat k chybám, kdy je slovo pravopisně napsáno podle pravidel, ale i tak je jeho použití ve větě špatné?

Tyto chyby můžeme klasifikovat jako překlady a tedy lexikologické chyby (anglické „*from*“ vs. „*form*“, neboli „*z*“ vs. „*forma*“), kognitivní či fonetické chyby (například anglické „*there*“ vs. „*their*“, v překladu „*tam*“ vs. „*jejich*“), další jsou syntaktické či gramatické prohřešky (špatné skloňování či časování slov), dále sémantické anomálie (jde o věty nebo slovní spojení, která respektují pravidla morfologie, gramatiky a syntaxe, ale nedávají smysl z hlediska logiky) a nakonec vynechání či přidání celých slov. Dále nesmíme zapomenout na chyby vzniklé nedodržením hranic slov (Kukich, 1992, s. 412). V češtině se jedná například o nesprávné psaní spřežek („*na krátko*“ vs. „*nakrátko*“).

V následujícím seznamu jsou uvedeny základní druhy přístupů k identifikaci těchto chyb:

- techniky založené na postupném uvolňování kritérií

Tyto systémy využívají převážně syntaktická pravidla (Kukich, 1992, s. 419). Zpracováváný text se nejdříve snaží analyzovat, identifikovat v něm věty a ty poté rozložit na jednotlivá slova. Následně se pokouší všem těmto slovům přiřadit syntaktickou funkci

v rámci jejich věty pomocí syntaktických pravidel. Cílem je tedy určit u každého slova zda se jedná například o podmět, přísudek, objekt atd. Pokud toto není možné, protože je věta napsána nesprávně, celý proces se opakuje. Tentokrát je ale určité syntaktické pravidlo při analýze vynecháno či není aplikováno tak striktně jako v předchozím cyklu. Toto se opakuje až do té doby, kdy je všem slovům přiřazena jejich funkce. Program, který využívá tuto techniku, může uživateli pomoci určit podstatu chyby, jelikož ví, jaká syntaktická pravidla musel vynechat či poupravit během kontroly (Mitton, 1996, s. 97).

- techniky založené na očekávání následného slova v závislosti na předcházejícím

Tyto systémy také rozkládají text na věty a slova pomocí syntaktických pravidel jako předchozí. Jejich odlišnost je v přístupu k identifikaci chyby. Jak syntaktický analyzátor (v anglickém originále označován jako „parser“) postupně zpracovává vstupní text, vytváří seznam slov, která očekává na následující pozici v textu v zásadě na jeho porozumění syntaktických a někdy i sémantických pravidel. Pokud se následující slovo na vstupu nenachází na tomto seznamu, systém předpokládá, že došlo k chybě (Kukich, 1992, s. 419).

- techniky založené na statistických jazykových modelech

Statistický jazykový model (v anglickém originále označován jako „statistical language model“ nebo zkratkou SML) je ve své podstatě tabulka odhadů pravděpodobností pro některá či veškerá slova jazyka udávající jejich šanci se vyskytovat v kontextu s ostatními slovy v jejich okolí (Kukich, 1992, s. 423). Proto například statistický jazykový model pracující se slovními trigramy (stejný princip jako u n-gramů uváděných u písmen – slovní n-gram tedy označuje  $n$  počet slov jdoucích po sobě) specifikuje pravděpodobnost výskytu slova v závislosti na předcházejících dvou slovech. Statistický jazykový model pracující se slovními spojeními či frázemi určuje šanci určitých slov se vyskytovat v blízkosti (například o pět slov v každém směru) dalšího lingvisticky příbuzného slova (Kukich, 1992, s. 424). Potřebná data k vytvoření těchto modelů se získávají z velkých vzorků textu čítajících až několik desítek milionů slov. Velikost seznamu možných slovních trigramů roste s třetí mocninou velikosti zdrojového textu, takže plnohodnotný model pro text čítající pět tisíc slov by měl dvacet pět trilionů vstupů (Kukich, 1992, s. 424). Toto názorně ilustruje náročnost těchto systémů. V dnešní době máme k dispozici značnou výpočetní sílu a kapacitu, i tak se ale jedná o těžké problémy. A to se tyto systémy zabývají zejména syntaxí, což teprve sémantikou. Ta vyžaduje analýzu na ještě

vyšší úrovni, přesahující jednotlivě věty a zahrnující i jejich logické vztahy a leckdy i vazby na obecný kontext.

- techniky pracující s pravděpodobností výskytu slov

Předcházející statistické jazykové modely pracují s pravděpodobností, existuje ale také ještě alternativní přístup. Ten opět rozloží text na jednotlivá slova a určí jejich syntaktické funkce a slovní druhy. Dále se ale nezabývá s pravděpodobností následného výskytu konkrétních slov, ale uvažuje vše obecněji. Pro jasnější ilustraci, pokud se na vstupu objeví anglická věta „The fly bit the goat.“ (Moucha kousla kozu.) je vše v pořádku, na úrovni slov ale dochází k určité mnohoznačnosti. Při analýze systém určí, že slova „fly“ a „bit“ mohou být slovesem i podstatným jménem (fly – moucha / létat; bit – minulý čas slovesa kousnout / bit jako jednotka velikosti informace) a že v případě slova „the“ se jedná o určitý člen. Ve statistické tabulce si vyhledá, že určitý člen je velice často následován podstatným jménem a naopak zřídka slovesem. V tomto případě tedy určí „fly“ jako podstatné jméno a „bit“ jako sloveso (Mitton, 1996, s. 98). Za použití tohoto principu lze například identifikovat při opravě nesmyslné spojení „the equalize“ (určitý člen „the“ a sloveso „vyrovnat“), kdy určitý člen následuje sloveso, což je statisticky velice nepravděpodobné.

- využití neuronových sítí

Jednou z nejposlednějších oblastí výzkumu ohledně oprav textu založených na kontextu mezi slovy je využití modelů neuronových sítí. Ty mají mnoho společného s již zmíněnými statistickými jazykovými modely. Také se pokoušejí zachytit vzájemné vztahy mezi slovy na základě jejich kontextu a u těchto relací se snaží vyjádřit jejich pravděpodobnost výskytu. Hlavním rozdílem oproti statistickým jazykovým modelům, které vyjadřují váhu očekávaného vztahu mezi slovy explicitně pomocí pravděpodobnosti, neuronové sítě toto implicitně reprezentují pomocí síly váhy spojení ve své struktuře. Dalším důležitým rozdílem je vysoká náročnost neuronových sítí na výpočetní výkon a paměť počítačů. Pro statistické jazykové modely je možné vytvořit seznam čítající stovky tisíců pravděpodobnostních vztahů mezi mnoha tisíci slovy a ten uložit na několika paměťových médiích. Naproti tomu, pro neuronové sítě je nutné uchovávat plně v paměti počítače tisíce buněk reprezentujících slova a stovky tisíců vah vyjadřujících sílu asociací mezi těmito slovy. Tyto nároky činí neuronové sítě bohužel nepraktické pro slovníky čítající více jak několik tisíc slov. Nicméně neuronové sítě by do budoucnosti mohly

pomoci zvýšit inteligenci systémů pracujících se vztahy mezi slovy na základě kontextu (Kukich, 1992, s. 428-9.).

### **2.4.2 Nabízení řešení**

Moderní prostředky na kontrolu textu pracují na interaktivní bázi. Prohledají zdrojový dokument a pokud nějaké slovo identifikují jako vadné, nabídnou uživateli možné opravy. Ten má dále několik možností, jak následně postupovat:

*Nahrazení.* Neznámé slovo (nebylo nalezeno ve slovníku) je bráno jako špatně napsané a bude nahrazeno správnou formou, jejíž zadání je vyžadováno po uživateli

*Nahrazení a zapamatování.* Neznámé slovo je nahrazeno jeho opravou zadanou uživatelem. Pokud bude příště slovo napsáno se stejnou chybou, bude nahrazeno touto opravou.

*Přijmutí / přeskočení.* Neznámé slovo je považováno za správné a ponecháno ve své podobě (Peterson, 1980, s. 679).

*Přijmutí / přeskočení a zapamatování.* Neznámé slovo je opět považováno za správné a ponecháno beze změny. Navíc bude toto slovo vždy považováno za korektní a již nebude po uživateli požadována jeho oprava (Peterson, 1980, s. 680).

Na základě jakých postupů ale dokáže program určit vyhovující kandidáty na opravu chyby a tím ulehčit uživateli celý proces?

#### **2.4.2.1 Techniky pro opravu chyby či chyb v rámci slova**

Problém řešící opravu slova se skládá z několika dílčích postupů. Nejprve je nutno identifikovat špatně napsané slovo (viz 2.4.1 Identifikace chyby), dále je potřeba vygenerovat seznam možných oprav a v poslední řadě je tento seznam seřazen (Kukich, 1992, s. 392.). Proces generování slov, která mohou sloužit jako oprava, často využívá slovníky nebo databáze n-gramů pro vyhledání potencionálních oprav. Proces seřazení kandidátních slov na opravu používá vztahy na základě lexikální podobnosti mezi špatně napsaným slovem a nabízenou opravou či odhad pravděpodobnosti že určité chybně napsané slovo mělo představovat jiné, správné. Některé techniky vynechávají třetí proces a nechávají tak seřazení a vybrání nejlepší opravy na uživateli. Další, jako například neuronové sítě či techniky pracující s pravděpodobnostmi, kombinují všechny tři procesy (identifikace, generování opravy, seřazení) do jednoho kroku vypočítáním pravděpodobnostního či podobnostního vztahu mezi slovem na vstupu (kontrolovaným slovem) a slovy ve slovníku. Pokud výsledné slovo ze slovníku, které se umístilo na nejvyšší pozici v seznamu (má nejvyšší hodnotu

pravděpodobnosti či podobnosti), není totožné s kontrolovaným slovem, je toto bráno jako chyba a poté je vypočítán seřazený seznam potencionálních oprav (Kukich, 1992, s. 393.).

Podle K. Kukich (Kukich, 1992) existuje šest základních postupů pro opravu chybně napsaného slova:

- techniky založené na minimálním počtu kroků nutných k opravení chyby

Jedněmi z nejvíce studovaných procesů na opravu textu jsou algoritmy vypočítávající minimální editační vzdálenost mezi špatně napsaným slovem a zápisem ve slovníku. Termín minimální editační vzdálenost (v anglickém originále „minimum edit distance“) je definován jako nejmenší počet úprav (nahrazení, vynechání, vložení písmena) nutných k transformaci jednoho slova ve druhé. Tato technika pracuje s předpokladem, že slovo, či slova, která jsou ve slovníku nalezena na základě vstupního špatně napsaného, budou jeho správným tvarem. Pro tyto algoritmy obecně platí, že vyžadují  $m$  porovnání mezi špatně napsaným slovem a slovníkem, kde  $m$  reprezentuje počet vstupů ve slovníku. Proto existuje několik postupů na zmenšení času nutného na vyhledání slova s minimální editační vzdáleností. Lze například vyjít z poznatku, že nejčastější chyby vznikají vynecháním jednoho písmena ze slova. Mor a Fraenkel tak docílili lepších vyhledávacích časů uchováváním každého slova ve slovníku  $x + 1$  krát, kde  $x$  reprezentuje délku slova, pokaždé s vynecháním jednoho písmena. Pro vyhledávání ve slovníku používali poté hashovací funkci. „Reverzní“ technika pro vyhledání minimální editační vzdálenosti také poskytuje rychlejší výsledky. Při reverzním přístupu je u špatně napsaného slova nejdříve vytvořen seznam všech jeho permutací na základě nahrazení / vynechání / vložení jednoho písmena. Poté se porovnáním tohoto seznamu se slovníkem zkoumá, zda permutací nevznikla korektní forma slova.

- techniky založené na klíči podobnosti

Záměrem těchto technik je převést slova na klíče tak, aby podobná slova měla podobné či dokonce identické klíče. Tím pádem klíč vypočítaný pro špatně napsané slovo slouží jako ukazatel na podobná slova uložená ve slovníku a tím i na kandidáty na opravu. Vyhledávání na tomto principu je velice rychlé, jelikož není nutné špatně napsané slovo přímo porovnávat s každým slovem ve slovníku. Jako příklad principu fungování může posloužit systém SOUNDEX z roku 1918 který byl mimo jiné použit v rezervačním systému pro aerolinky z roku 1962. Převádí slova na klíče sestávající z jejich prvního

písmena následovaného skupinou číslic. Čísla jsou přiřazována pomocí následujících pravidel:

A, E, I, O, U, H, W, Y ... 0

B, F, P, V ... 1

C, G, J, K, Q, S, X, Z ... 2

D, T ... 3

L ... 4

M, N ... 5

R ... 6

Nuly jsou vynechávány a opakované číslice jsou nahrazeny pouze jednou. Slovo „Bush“ pak získá klíč B020 neboli B2 a jeho nesprávná forma „Busch“ klíč B0220 – také B2.

- techniky založené na pravidlech transformace chybně napsaného slova

Tyto techniky jsou algoritmy snažící se reprezentovat znalosti o často se vyskytujících vzorech v chybných tvarech slov ve formě pravidel, která umožní transformaci chybně napsaného slova na jeho korektní podobu. Proces na vytvoření seznamu možných oprav se sestává z užití všech aplikovatelných transformačních pravidel na chybné slovo a následné uchování všech vzniklých podob, které se vyskytují i ve slovníku.

- techniky založené na n-gramech

N-gramy písmen (skupina  $n$  sousedících písmen), obsahující například trigramy (tři sousední písmena), bigramy (dvě sousední písmena) či unigramy (jedno písmeno), jsou využívány mnoha způsoby při optickém rozeznávání znaků (OCR – optical character recognition) a opravách pravopisu. V systémech oprav slouží jako přístupové klíče odkazující ve slovníku na možná kandidátní slova pro opravu a také slouží jako lexikální vlastnost slov umožňující vypočítat podobnost slov.

- techniky pracující s pravděpodobnostmi

Tyto techniky používají dva druhy pravděpodobnosti: tranzitivní / přechodové pravděpodobnosti a záměnné pravděpodobnosti. Tranzitivní či přechodové pravděpodobnosti reprezentují pravděpodobnost, že dané písmeno či skupina písmen budou následovány určitým písmenem, či skupinou. Tyto pravděpodobnosti nejsou obecně platné pro všechny jazyky, pro každý jsou specifické. Data nezbytná k vytvoření těchto odhadů pravděpodobností se získávají ze statistik výskytů n-gramů ve velkých vzorcích textu. Záměnné pravděpodobnosti odhadují, jak často se dané písmeno

zaměňuje s jiným určitým písmenem. Tyto odhady jsou závislé na zdroji, proto například při opravách textu vzniklého při optickém rozpoznávání textů je pro každý stroj vykonávající tuto funkci generována specifická hodnota pravděpodobností. Ty záleží na také na fontu a velikosti zpracovávaného textu. Záměnné pravděpodobnosti založené na chybách člověka se jednoduše nazývají chybové pravděpodobnosti. Ty byly získány vyhodnocením velkého objemu textu psaného člověkem obsahujícího překlepy a další pravopisné chyby. Výzkum v oblasti rozpoznávání textu ukázal, že tyto pravděpodobnosti odhady samy o sobě nedokážou poskytnout uspokojující výsledky při opravě chyb, nicméně kombinace těchto odhadů s technikami porovnávání se slovy ve slovníku přináší již daleko lepší výsledky.

- **techniky pracující s neuronovými sítěmi**

Neuronové sítě jsou nadějnými kandidáty pro využití při opravách pravopisu pro jim vlastní schopnost dojít k výsledku i za využití nekompletních či poškozených dat na jejich vstupu. Navíc mohou být také trénovány na reálných pravopisných chybách (za použití slov tyto chyby obsahující; poznámka pro upřesnění překladu z originálu) a tím mají potenciál adaptovat se na specifický výskyt chyb u různých skupin uživatelů. Lze si dokonce představit čip obsahující neurální síť umístěný v osobních stanicích, který by se neustále průběžně adaptoval na specifika chyb jejich uživatelů. Využití neuronových sítí je velice zajímavé pro jejich dovednost vypořádat se s neúplnými a poškozenými daty a jejich schopnost se učit. Ta je ale i jejich nedostatkem – před použitím musí nejdříve neuronové sítě projít procesem učení, který vytvoří jejich strukturu. To je velice časově náročné, leckdy vyžadující i stovky hodin práce.

## **2.5 Představení textových procesorů z hlediska systémů oprav textu**

Většina běžných uživatelů počítačů přišla nejspíše do styku převážně s textovým procesorem MS Office Word díky masovému rozšíření jeho mateřského operačního systému. Existují ale i alternativy k tomuto programu a ty mohou k problematice oprav textu přistupovat jiným způsobem. Proto následující část tyto programy krátce představuje a nastiňuje jejich funkčnost z hlediska automatických oprav textu.

### **2.5.1 OpenOffice.org Writer**

Jak již napovídá nadpis, textový procesor Writer je součástí balíčku kancelářských aplikací OpenOffice.org. Jedná se o open-source alternativu ke komerčnímu MS Office Word, jejíž výhodou je její cena – je ke stažení zcela zdarma. Na domovských stránkách produktu se

uvádí, že prošel více jak dvacetiletým vývojem. Jeho open-source charakter umožňuje komukoliv nahlásit chybu, vyžádat si novou funkci, nebo rozšířit funkcionalitu programu (Why OpenOffice.org, [cit. 2012-03-15]).

V dokumentu „General Differences in Use between OpenOffice.org and Microsoft Office“ (Obecné rozdíly v užití OpenOffice.org a Microsoft Office) je napsáno, že Writer má oproti Microsoft Office funkci „auto-complete“ (automatické dokončení slova či fráze). Když uživatel píše, tak se textový procesor kontinuálně snaží předpovědět, jaké by mohlo být dokončení právě psaného slova. Nastavení této funkce můžeme najít v: **Nástroje > Nastavení automatických oprav > Dokončování slov** (General Differences in Use between OpenOffice.org and Microsoft Office, [cit. 2012-03-21], s. 1).

Při otevření nového dokumentu je seznam pro automatické dokončování prázdný. Nelze jej ihned naplnit nebo nahrát seznam z externího souboru. Seznam se vytváří sám při psaní dokumentu, program vyhodnocuje četnost výskytu slov v dokumentu a poté přidává ta nejčastější do seznamu. Funkci automatického dokončování tak lze aktivně využívat až po napsání určité části dokumentu. Zaškrtnutím volby „Při zavření dokumentu odstranit slova z něj ze seznamu“ (nejde o chybu přepisu, takto je provedena česká lokalizace ve verzi 3.3.0) je možné povolit sdílení seznamu mezi právě otevřenými dokumenty, nicméně po zavření programu OpenOffice.org dojde i tak ke smazání seznamu. (čerpáno z vlastní zkušenosti a z nápovědy programu OpenOffice.org)

Pokud dojde při napsání prvních tří písmen nového slova ke shodě se slovem uloženým v seznamu automatického dokončování, textový procesor nabídne toto slovo k dokončení. To lze speciální klávesou potvrdit, v základním nastavení se jedná o klávesu Enter (čerpáno z vlastní zkušenosti a z nápovědy programu OpenOffice.org).

Podobného efektu lze nicméně docílit i v MS Office Word právě použitím automatických oprav. Do jejich seznamu lze manuálně přidat slovo, jehož psaní si chceme ulehčit a přiřadit mu „špatně psanou formu“, kterou ho poté vyvoláme. Pokud například často používáme adresu sídla naší firmy, stačí do sloupce „Nahrazovat:“ vyplnit třeba řetězec „(adr)“ a do sloupce „Za:“ uvést onu adresu. Poté pokaždé, když v dokumentu MS Office Word napíšeme „(adr)“ bude toto automaticky nahrazeno adresou. Výhodou programu OpenOffice.org Writer zůstává fakt, že jeho seznam je vytvářen automaticky kontinuálně bez nutnosti našeho vlastního zásahu.



Dalším rozdílem oproti Microsoft Office na poli automatických oprav je absence gramatické kontroly (zde se jedná o překlad z anglického „grammar checking“, gramatika je zde chápána jako skloňování slov a jejich syntaktické, morfologické a sémantické vztahy) v OpenOffice.org Writer (General Differences in Use between OpenOffice.org and Microsoft Office, [cit. 2012-03-21], s. 16).

### **2.5.2 Google Documents (docs.google.com)**

Google Docs je balíček produktů který umožňuje uživateli vytvořit různé druhy online dokumentů, pracovat na nich v reálném čase spolu s více lidmi a ukládat dokumenty a další soubory online. A to vše zcela zdarma. S připojením na internet je možné přistupovat k dokumentům a souborům z jakéhokoliv počítače kdekoli na světě. Určité úkony lze dokonce provádět i v offline režimu (An overview of Google Docs - Google Docs Help, [cit. 2012-03-24]). A právě součástí Google Docs je i online textový procesor Google Documents.

V programu funguje standardní kontrola pravopisu, která nachází slova se špatným pravopisem a nabízí možnou opravu. Slova, u kterých je nabízena alternativní podoba, jsou zvýrazněna červeným podtržením. Opravy se zobrazí jednoduše pravým kliknutím na zvýrazněné slovo. Program nabízí také automatické opravy a rozšiřování slovníku o další slova (Spell check and automatic corrections - Google Docs Help, . [cit. 2012-03-24]). Má tedy veškeré funkce, na které jsme zvyklí u offline verzí textových procesorů.

### **3 Praktická část**

Praktická část práce se věnuje vlastnímu výzkumu. Ten byl zaměřen na zkoumání souvislostí využití prostředků automatických oprav a oprav pravopisu a gramatiky v textu a gramatického projevu u zvoleného testovaného vzorku.

#### **3.1 Vlastní výzkum**

Výzkum byl prováděn s laskavým svolením vyučujícího, pana magistra Vargy, a vedení školy Gymnázia Plasy v hodinách výuky ICT tříd kvinta osmiletého gymnázia a 1. ročníku čtyřletého gymnázia (třídy ve školním roce 2011/2012). Seznámení žáků s výzkumem a jeho následná realizace zabrala vždy jednu celou vyučovací hodinu.

Výzkum byl prováděn čistě anonymně, není možné jakkoliv zpětně přiřadit konkrétního žáka k datům získaným v průběhu výzkumu. Jedinými vyžadovanými osobními údaji bylo pohlaví a věk.

Pro realizaci části výzkumu, která je závislá na práci v textovém procesoru, byl zvolen Microsoft Office Word 2007. Ostatní produkty jsou buď rozšířeny v mnohem menší míře a jejich dopad na gramatiku mládeže lze předpokládat jako minimální, nebo tyto produkty „dožívají“ a lze předpokládat, že vliv jejich systémů oprav bude stále nižší (například Microsoft Office Word 2003).

##### **3.1.1 Vymezení cíle a hypotéz výzkumu**

Cílem výzkumu je zjistit, zda existuje vazba mezi využitím prostředků automatických oprav a oprav pravopisu a gramatiky v prostředí textového procesoru Microsoft Office Word 2007 a úrovni gramatického projevu mládeže. Dalším cílem je zhodnocení rozložení chyb v rukou psaném textu z hlediska jejich rozdělení na chyby rozpoznávané procesorem MS Office Word 2007 a chyby tímto procesorem nerozpoznávané. Tím, že procesor upozorňuje uživatele na chyby, s ním interaguje a je možné, že ovlivňuje i jeho gramatický projev a učí ho gramatice. Relativní počet rozpoznávaných chyb v textu psaném rukou by se tak mohl lišit od relativního počtu chyb nerozpoznávaných, mohl by být například nižší.

Výzkum se také zabývá pohledem účastníků na obecné využívání prostředků automatických oprav textu, míru jejich využití a vnímání důležitosti gramatiky jako takové. Na základě těchto kritérií jsou účastníci rozděleni na čtyři skupiny: využívající automatické opravy, nevyužívající automatické opravy, příkládající gramatice důležitost a nepřikládající gramatice

důležitost. Dílčím cílem výzkumu pak je odhalit možnou souvislost mezi úrovní gramatického projevu účastníku a jejich příslušnosti do jedné ze čtyř dříve zmíněných skupin.

Pro tento výzkum byly zvoleny následující hypotézy. Hlavním úkolem hypotéz je odhalit a případně potvrdit možný vliv zvoleného textového procesoru na gramatickou úroveň projevu účastníků výzkumu.

H1: *V textu psaném rukou u žáků tříd kvinta a 1. ročník Gymnázia Plasy je počet chyb rozpoznávaných programem Microsoft Office Word 2007 nižší než počet chyb tímto programem nerozpoznávaných.*

H2: *Počet chyb nerozpoznávaných textovým procesorem Microsoft Office Word 2007 v pokusném textu napsaném pomocí tohoto procesoru je vyšší než počet stejného druhu chyb ve stejném textu psaném rukou.*

Žáci tříd kvinta a 1. ročník Gymnázia Plasy jsou na základě odpovědí v dotazníku rozděleni na skupinu aktivně využívající prostředky automatických oprav textu (skupina A) a na skupinu tyto prostředky méně využívající (skupina B).

H3: *Počet chyb v diktátu psaném rukou je u skupiny A nižší než počet chyb u skupiny B.*

Žáci tříd kvinta a 1. ročník Gymnázia Plasy jsou na základě odpovědí v dotazníku rozděleni na skupinu přisuzující gramatice důležitost (skupina C) a na skupinu vnímající gramatiku jako méně důležitou (skupina D).

H4: *Počet chyb v diktátu psaném rukou je u skupiny C nižší než počet chyb u skupiny D.*

### **3.1.2 Zdůvodnění výběru vzorku a jeho charakteristika a popis**

Výběr škol, bohužel, neprobíhal standardním náhodným způsobem. Původní záměr vybrat náhodně alespoň 3 školy různého typu a zaměření narazil na objektivní překážky. Jednalo se o organizační záležitosti, neochotu škol poskytnout prostor pro realizaci výzkumu či nedostatečné vybavení PC učeben (pro výzkum nevhodný software). Nakonec byla zvolena jediná škola, která byla schopna poskytnout čas i prostor pro výzkum a nabídnout odpovídající vybavení. Jako částečná kompenzace nenaplnění původního plánu, byly zapojeny do výzkumu dvě třídy zvolené školy. Jako zkoumaný vzorek byli tedy určeni žáci Gymnázia Plasy docházející do tříd kvinta osmiletého gymnázia a 1. ročník čtyřletého gymnázia.

Třídy kvinta a 1. ročník byly zvoleny, jelikož se jedná o první třídy na osmiletém a čtyřletém gymnáziu, které již mají ukončené základní vzdělání, a tudíž by jejich žáci měli být celkově

obeznání s českou gramatikou. Mělo by tím být eliminováno nebezpečí, že by mohli účastníci výzkumu v testu chybovat kvůli neznalosti gramatických pravidel z důvodu nedokončeného vzdělání v této oblasti. Výběr těchto dvou tříd byl také přínosný z organizačních důvodů, umožnil totiž provádět výzkum v rámci jednoho dne.

Výhodou výběru Gymnázia Plasy pro konání výzkumu je také fakt, že se tato škola skládá z osmiletého a čtyřletého gymnázia. Nabízí se tak také možnost porovnat úroveň gramatického projevu žáků paralelních tříd těchto dvou typů školy. Nejde o primární záměr tohoto výzkumu, nicméně byla by škoda nevyužít možnosti pro toto srovnání, když jsou již k dispozici potřebná data.

Výzkumu se zúčastnilo celkem 44 žáků. Jejich rozdělení do tříd je následující:

	<b>muži</b>	<b>ženy</b>	<b>celkem</b>
<b>1. ročník</b>	9	8	17
<b>kvinta</b>	12	15	27
<b>celkem</b>	<b>21</b>	<b>23</b>	<b>44</b>

**Tabulka 1: Rozdělení respondentů podle tříd a pohlaví**

Jak je vidět z Tabulky 1, rozložení mužů a žen je ve zkoumaném vzorku prakticky stejné, žen je pouze o dvě více, vyjádřeno procenty tedy 47,7% mužů a 52,3% žen. To nabízí prakticky ideální podmínky pro porovnání výsledků výzkumu i z hlediska pohlaví respondentů, což může obohatit původně zamýšlené výstupy z výzkumu o zajímavé poznatky.

Průměrný věk žáků zkoumaného vzorku je šestnáct let, což odpovídá typickému věku žáka zvolených tříd. Nejnížší věk respondentů je patnáct let a nejvyšší sedmnáct, nejčastěji uváděný věk činí šestnáct let.

### **3.1.3 Plán výzkumu**

Pro získání potřebných dat pro realizaci výzkumu je využito v první fázi dotazníku. Jeho podoba je k nahlédnutí v sekci 5. Přílohy této práce. Dotazník se skládá celkem z patnácti položek. První čtyři, podle pořadí v dotazníku, slouží k identifikaci jednotlivých respondentů pomocí identifikačního čísla a k poskytnutí dat nutných k rozdělení respondentů do skupin podle pohlaví a třídy, do které dochází. Dalších jedenáct položek se týká již samotného výzkumu. Ty si dávají za cíl zjistit, do jaké míry je pro respondenty důležitá jejich úroveň gramatického projevu a jak často a jakým způsobem využívají prostředky automatických oprav textu.

U položek využívaných k identifikaci respondentů (v dotazníku v části „Mé údaje“) lze vybírat ze dvou možností u volby pohlaví, ze dvou možností u volby příslušnosti ke třídě (kvinta či 1. ročník), další položkou je tvořená odpověď zjišťující věk žáka a poslední položka určuje jeho identifikační číslo v rámci celého výzkumu.

V druhé části dotazníku věnující se výzkumu jsou u každé položky nabízeny minimálně tři odpovědi a maximálně čtyři. Respondent musí zvolit pouze jednu z nabízených odpovědí. U šesti položek je ještě navíc tvořená odpověď na otázku „Proč“, která má za úkol odhalit hlavní důvody žáků pro zvolení určité odpovědi.

Pro realizaci dotazníku je využito prostředků nabízených v rámci online aplikace Google Documents. Ta nabízí možnost jednoduchého vytvoření elektronických dotazníků a jejich následné rozeslání e-mailem či prezentování na internetu. Jednotlivé odpovědi jsou okamžité po vyplnění a odeslání dotazníku uloženy na serveru Google Documents. Po získání dostatečného počtu odpovědí lze přijímání dalších zablokovat a obdržená data lze exportovat do přehledné tabulky MS Office Excel. Toto řešení ušetří v první fázi vyhodnocování výzkumu mnoho času, získaná data jsou ihned roztríděná k dispozici. Další podstatnou výhodou je možnost nastavit veškeré odpovědi v dotazníku jako povinné. Dotazník není možné před vyplněním všech takto nastavených položek odeslat, sám si vyžádá vyplnění těch zbývajících. Odpadá tak nutnost sledovat žáky během práce na dotazníku, stačí si pouze na konci zkontrolovat, zda bylo do online databáze uloženo tolik vstupů, kolik je respondentů. Tak je zaručeno, že každý z nich plně odpověděl na otázky.

Po uzamknutí dotazníku je nutné data nějakým způsobem interpretovat. U každé položky je spočítána četnost jednotlivých odpovědí, ta je ve vyhodnocení výzkumu vyjádřena absolutně a relativně (číselnou formou a v procentech).

Jedním z cílů dotazníku je rozdělit respondenty na čtyři skupiny podle jejich odpovědí. První až sedmá otázka podle pořadí v dotazníku slouží k rozdělení respondentů podle důležitosti, kterou přisuzují správné gramatice. Osmá až jedenáctá otázka pak rozděluje respondenty podle míry využití automatických oprav textu. Aby toto bylo možné, je nutné vyjádřit jejich odpovědi způsobem, který by umožnil přímé srovnání. Každá z odpovědí na položku dotazníku je proto bodově ohodnocena. Odpovědi, které ukazují na fakt, že respondenti využívají prostředky automatických kontrol textu a že přikládají gramatice důležitost, jsou ohodnoceny více body, než odpovědi opačné. Následující tabulka udává rozložení bodů pro jednotlivé odpovědi u položek dotazníku.

Pokud je vám vrácen test z jiného předmětu než z Čj a je v něm opravena gramatická chyba, znamená to pro vás:			
vítaný přínos	chybu ignoruji	gramatické chyby nám opravují pouze učitelé v rámci předmětu Čj	
4	0	2	
Při osobní neveřejné komunikaci s vašimi přáteli (např. pomocí sms, icq, chatu) máte podezření, že jste udělali gramatickou chybu (nejde o překlep), co uděláte?			
přesvědčím se, zda je vše správně a až poté pošlu zprávu	neřeším to, při komunikaci s přáteli na gramatice nezáleží	tyto zprávy ani nekontroluji a ihned odesílám	
4	0	0	
Pokud najdete na internetu (např. v článku, diskuzi) gramatickou chybu:			
autora na ni upozorním	neřeším to	chyb si nevšímám	
4	2	0	
Pokud píšete příspěvek, který může kdokoli na internetu vidět (např. diskuze, blog, článek, návštěvní kniha) a který není možné spojit s vaší osobou (zůstáváte anonymní), je pro vás gramatika:			
nepodstatná	důležitá	nad gramatikou při psaní nepřemýšlím	
0	4	2	
Je pro vás důležitá gramatika při odevzdávání úkolu v jiném předmětu než v Čj?			
ano	nepřemýšlím nad tím		ne
4	2		0
Pokud vás někdo upozorní na gramatickou chybu, kterou jste napsali například v diskuzi nebo na facebooku:			
pokusím se ji příště neopakovat	ignoruji to	ohradím se	nikdo mne neopravuje
4	0	2	2
Pokud píšete příspěvek, který může kdokoli na internetu vidět (např. diskuze, blog, článek, návštěvní kniha) a který je možné spojit s vaší osobou (nejste anonymní), je pro vás gramatika:			
nepodstatná	důležitá	nad gramatikou při psaní nepřemýšlím	
0	4	2	
Pokud vám automatická oprava zvýrazní špatně napsané slovo:			
vyberu z nabídky opravy	zamyslím se a snažím se slovo sám opravit	ignoruji to	automatické opravy nepoužívám, mám je vypnuté
2	4	0	0
Automatické dokončování slov při psaní sms zpráv, např. T9:			
používám a doplňuji nová slova do slovníku	používám, ale slovník neaktualizuji	nepoužívám	
4	2	0	
Automatické opravy (například v MS Office Word) máte:			
zapnuté	vypnuté	nevěnuji jim velkou pozornost	
4	0	2	
Používáte automatické opravy a / nebo dokončování slov:			
pro usnadnění práce	jako prostředek ke kontrole gramatiky	pro opravu a jako pomůcku při učení se gramatiky	tyto prostředky nepoužívám
2	4	4	0

Tabulka 2 Bodování odpovědí

Následně lze u každého respondenta určit pomocí tohoto klíče souhrnnou hodnotu, která určuje jeho pořadí vůči ostatním v závislosti na vybraných odpovědích. Čím více bodů žák získá, tím více by měl používat prostředky automatických oprav a měl by také více dbát na správný gramatický projev. Respondenti jsou rozděleni na čtyři skupiny podle počtu získaných bodů, kde dělicím kritériem je střední hodnota (medián) získaná ze souhrnných bodů všech respondentů. Dále lze určit, zda se projeví nějaké rozdíly v odpovědích v závislosti na pohlaví a třídě respondentů.

V druhé fázi je s žáky psán diktát, a to celkem dvakrát ve stejném zadání. Podoba diktátu je k nahlédnutí v sekci 5. Přílohy této práce. Při samotném provádění výzkumu budou diktáty předcházet dotazník, zde je uvádím jako druhou fázi podle pořadí při vyhodnocování výzkumu, při zpracování diktátů budou totiž používány některé informace získané z dotazníku.

Žáci nejdříve napíší diktát rukou na předem připravené linkované papíry. Jsou požádáni, aby nejdříve poznamenali svou třídu a také identifikační číslo, které jim je přiděleno a shoduje se s číslem použitým v druhém diktátu a v dotazníku. Po dopsání diktátu všemi žáky je diktát ihned vybrán bez možnosti kontroly, aby se předcházelo vzájemnému razení. Poté ihned následuje identický diktát, tentokrát ale psaný v prostředí Microsoft Office Word 2007. Žáci opět do dokumentů poznamenají svá identifikační čísla a třídy. Po dopsání je žákům oznámeno, že mají minutu na kontrolu a mohou podle uvážení použít kontrolu pravopisu a gramatiky. Poté jsou soubory uloženy do společného adresáře a zkopírovány na flash disk.

Při vyhodnocování diktátů jsou nejdříve oba diktáty přiřazeny k jednotlivým pisatelům. Každý diktát je opraven a je zaznamenán celkový počet chyb. U obou diktátů jsou následně chyby rozděleny na chyby rozpoznávané kontrolou pravopisu a gramatiky aplikace Microsoft Office Word 2007 a na chyby nerozpoznávané. Jejich počet je také zaznamenán. Rozpoznávané chyby jsou například: pravopisné chyby, překlipy. Mezi nerozpoznávané patří shoda podmětu s přísudkem, psaní spřežek (na krátko vs. nakrátko), psaní s / z (sběh vs. zběh) atd. Po získání těchto dat je možné sestavit tabulku, která udává identifikační číslo žáka, jeho věk, pohlaví, třídu, celkový počet bodů získaný v dotazníku, počet rozpoznávaných a nerozpoznávaných chyb napsaných v diktátu na PC a počet rozpoznávaných a nerozpoznávaných chyb napsaných v diktátu rukou. Z této tabulky je poté možno dále vycházet.

V konečné části vyhodnocení diktátů je porovnáváno několik hodnot:

- počet chyb nerozpoznávaných kontrolou pravopisu a gramatiky aplikace Microsoft Office Word 2007 vůči chybám rozpoznávaným a to v rukou psaných diktátech
- počet nerozpoznávaných chyb psaných v diktátu na PC vůči počtu nerozpoznávaných chyb psaných v diktátu rukou
- počet všech chyb napsaných v diktátu na PC vůči počtu všech chyb napsaných v diktátu rukou

Tyto hodnoty jsou pro každého žáka vypočítány zvlášť. Dále lze vypočítat jejich průměrnou hodnotu, či je třídit podle pohlaví či třídy jednotlivých žáků.

### **3.1.4 Postup při realizaci výzkumu**

Žáci byli nejdříve krátce seznámeni s výzkumem. Nebyly jim sděleny žádné podrobnosti, pouze že bude sloužit k vypracování bakalářské práce. Poté byli ujištěni, že veškeré získané informace jsou čistě anonymní. Žákům bylo sděleno, že se výzkum skládá ze dvou totožných diktátů a z dotazníku. Bylo jim vysvětleno, jak bude psaní diktátů probíhat. Byli také požádáni, aby psali podle svého nejlepšího uvážení, snažili se dosáhnout nejlepších výsledků, ale zároveň nepodváděli a neradili se.

Diktáty byly přednášeny autorem této bakalářské práce. Při diktování byla snaha o jasnou artikulaci a srozumitelný přednes. Po každé větě byla dána časová rezerva pro revizi a během diktování byla průběžně prováděna kontrola, zda všichni zúčastnění všemu rozumí. Žáci tak nebyli pod tlakem a bylo sníženo riziko, že by psali chyby kvůli časové tísní.

Po napsání diktátů byl žákům předložen elektronický dotazník. Bylo jim vysvětleno, že veškeré položky jsou povinné a že dotazník si jejich vyplnění kontroluje. Žákům byla dána možnost ptát se, pokud by čemukoliv v dotazníku nerozuměli.

### **3.1.5 Kontrola odpovědí**

Nedílnou součástí každého výzkumu musí být i kontrola nashromážděných dat před provedením samotného vyhodnocení výzkumu. Ten by měl počítat a operovat pouze s relevantními informacemi které odpovídají kritériím určeným v popisu tohoto výzkumu. Jinak mohou být výsledky ovlivněny nesprávnými či záměrně zavádějícími odpověďmi. Tento výzkum se naštěstí nesetkal s takovýmito obtížemi, nicméně dva účastníci, a tím i jejich odpovědi, museli být vyřazeni.



Prvním je žena, která bohužel přišla později do hodiny a tím pádem neměla možnost napsat celý diktát. Tato situace by mohla být vyřešena pozdějším dodatečným dopsáním diktátu, nicméně tento postup nakonec nebyl zvolen a její odpovědi jsou vyřazeny. Dá se předpokládat, že žáci po hodině o výzkumu diskutovali, proto by jí později napsaný diktát nebyl již relevantní – její gramatický projev by již nebyl autentický, protože již byla předem seznámena s obsahem diktátu a i samotného výzkumu jako celku.

Druhým vyřazeným byl muž, jehož rukou psaný diktát obsahuje velké množství chyb a vykazuje velice podivný a zajímavý trend – naprosto ignoruje tečky a mezery mezi větami. Místo teček používá čárky, za kterými dokonce chybí i mezera. Celý diktát je tak v podstatě jedna dlouhá věta. V jednom případě je dokonce na konci diktované věty správně napsán otazník, i tak ale za ním opět následuje čárka. Nicméně tento trend se neopakuje v diktátu vzápětí napsaném tímto mužem na počítači. Ten je již v naprostém pořádku. Rozdíl mezi počtem chyb napsaných v obou diktátech je tak velice značný a mohl by neadekvátně ovlivnit výsledky výzkumu. Je těžké uvěřit, že by se jednalo o běžný způsob psaní textu rukou, který by tento muž pravidelně používal, je tedy možné, že se jednalo o záměr. Podoba tohoto rukou psaného diktátu je k nahlédnutí v části 5. Přílohy pod označením Obrázek 1 Přepis diktátu vyřazeného z výzkumu. Diktát byl na požádání autora bakalářské práce přepsán druhou osobou, aby nebylo možné identifikovat pisatele diktátu na základě rukopisu a tím byla zachována jeho anonymita.

Konečný počet vyhodnocovaných účastníků výzkumu je tedy 42. Jejich rozdělení do tříd podle pohlaví je následující:

	<b>muži</b>	<b>ženy</b>	<b>celkem</b>
<b>1. ročník</b>	9	8	17
<b>kvinta</b>	11	14	25
<b>celkem</b>	<b>20</b>	<b>22</b>	<b>42</b>

**Tabulka 3 Konečný počet vyhodnocovaných účastníků výzkumu**

### 3.1.6 Výsledky

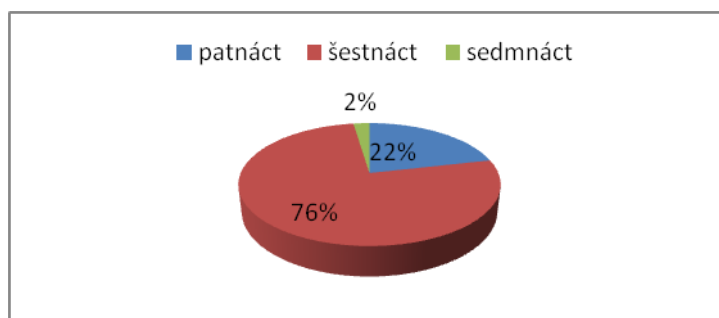
Výsledky dotazníku jsou zpracovány do tabulek a také grafů pro větší přehlednost. Jednotlivé položky jsou řazeny podle pořadí, jak se objevují v dotazníku. Vždy je uvedena absolutní četnost (počty odpovědí) a relativní četnost (počty v %). Nejčastější odpověď je v tabulce vždy tučně zvýrazněna pro snadnější identifikaci.

#### Položka II

*Prosím zadejte věk formou čísla.*

	relativní četnost	absolutní četnost
patnáct	21,43%	9
<b>šestnáct</b>	<b>76,19%</b>	<b>32</b>
sedmnáct	2,38%	1

Tabulka 4 Rozdělení podle věku



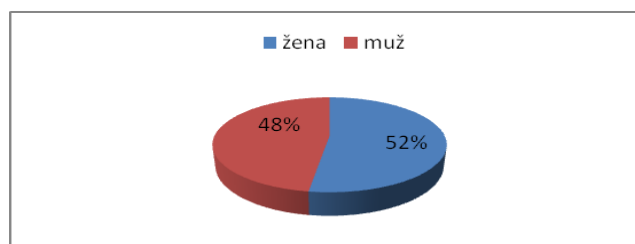
Z celkového počtu 42 respondentů dosáhlo 32 šestnácti let (76%), 9 patnácti let (21%) a 1 sedmnácti let (2%).

#### Položka III

*Pohlaví.*

	relativní četnost	absolutní četnost
<b>žena</b>	<b>52,38%</b>	<b>22</b>
muž	47,62%	20

Tabulka 5 Rozdělení podle pohlaví



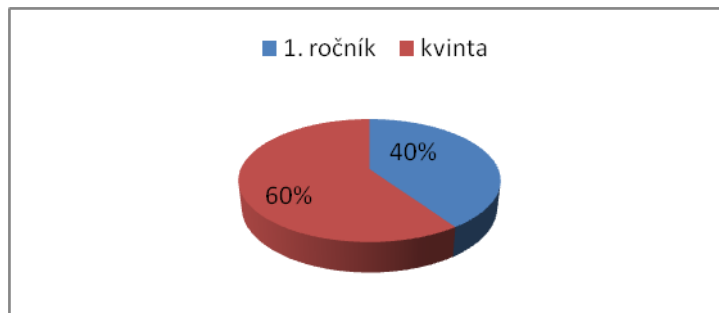
Z celkového počtu 42 respondentů bylo 22 žen (52 %) a 20 mužů (48%).

## Položka IV

*Třída.*

	relativní četnost	absolutní četnost
1. ročník	40,48%	17
kvinta	59,52%	25

Tabulka 6 Rozdělení podle tříd



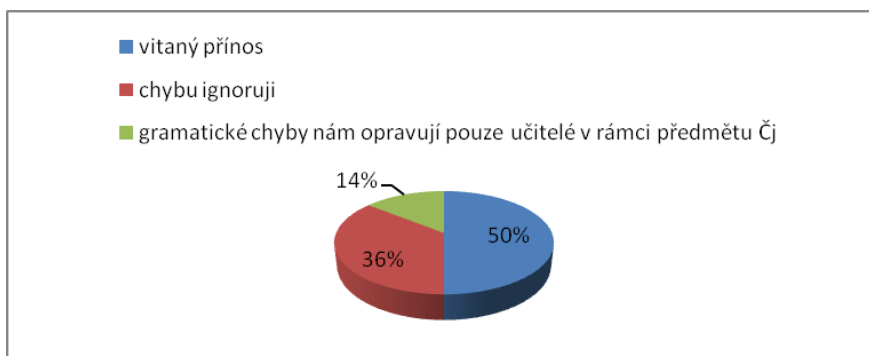
Z celkového počtu 42 respondentů docházelo 17 do třídy 1. ročník (40%) a 25 do třídy kvinta (60%).

## Položka V

*Pokud je vám vrácen test z jiného předmětu než z Čj a je v něm opravena gramatická chyba, znamená to pro vás:*

	relativní četnost	absolutní četnost
vítaný přínos	50,00%	21
chybu ignoruji	35,71%	15
gramatické chyby nám opravují pouze učitelé v rámci předmětu Čj	14,29%	6

Tabulka 7



Pokud je respondentům (celkový počet 42) vrácen test z jiného než z Čj a je v něm opravena gramatická chyba 21 z nich to vidí jako vítaný přínos (50%), 15 chybu ignoruje (36%) a 6 uvedlo, že gramatické chyby jsou jim opravovány pouze učiteli v rámci předmětu Čj (14%). Jako důvod pro výběr možnosti „vítaný přínos“ byl nejčastěji (v 8 případech z 21) uveden

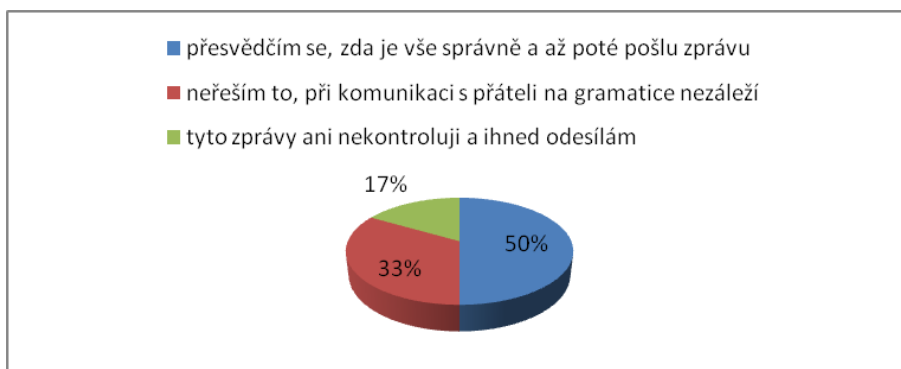
fakt, že se respondenti mohou chybu uvědomit a zapamatovat. Ve dvou případech uvedli, že oprava gramatické chyby mimo test z Čj je pro ně známkou toho, že učitel test opravdu četl a proto si i chyby všiml. U možnosti „chybu ignoruji“ nejčastěji (ve 12 případech z 15) jako důvod pro výběr uvedli, že se nejedná o Čj jazyk a že chyby neovlivní jejich klasifikaci. Respondenti, kteří si myslí, že jsou jim chyby opravovány pouze v rámci Čj, uvedli, že to ostatní učitele nejspíše nezajímá.

## Položka VI

*Při osobní neveřejné komunikaci s vašimi přáteli (např. pomocí sms, icq, chatu) máte podezření, že jste udělali gramatickou chybu (nejde o překlep), co uděláte?*

	relativní četnost	absolutní četnost
<b>přesvědčím se, zda je vše správně a až poté pošlu zprávu</b>	<b>50,00%</b>	<b>21</b>
neřeším to, při komunikaci s přáteli na gramatice nezáleží	33,33%	14
tyto zprávy ani nekontroluji a ihned odesílám	16,67%	7

Tabulka 8



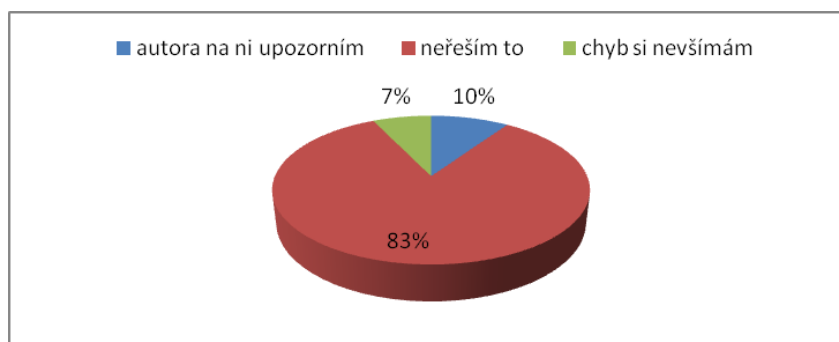
Z celkového počtu 42 respondentů 21 (50%) uvádí, že pokud mají při osobní neveřejné komunikaci s přáteli podezření, že udělali gramatickou chybu, raději se přesvědčí, zda je vše napsáno správně a až poté zprávu odešlou. Čtrnáct respondentů (33%) by toto neřešilo, jelikož při komunikaci s přáteli na gramatice nezáleží. Zbylých 7 (17%) tyto zprávy ani nekontroluje a ihned odesílá.

## Položka VII

*Pokud najdete na internetu (např. v článku, diskuzi) gramatickou chybu:*

	relativní četnost	absolutní četnost
autora na ni upozorním	9,52%	4
<b>neřeším to</b>	<b>83,33%</b>	<b>35</b>
chyb si nevšímám	7,14%	3

Tabulka 9



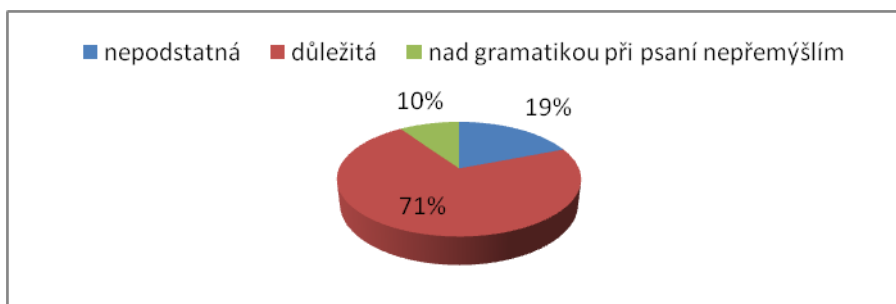
Z celkového počtu 42 respondentů 35 uvádí (83%), že pokud by na internetu (např. v článku, diskuzi) našli gramatickou chybu, nijak by toto neřešili. Čtyři respondenti (10%) by na chybu autora upozornili a zbylí 3 (7%) si chyb na internetu vůbec nevšímá. Nejčastější (v 9 případech z 35) důvod pro výběr možnosti „neřeším to“ byl, že se jedná o problém autora a ne respondenta. Dále žáci například uváděli, že by autor jejich upozornění ignoroval, nebo nechtěli autora svými připomínkami obtěžovat. U výběru „chyb si nevšímám“ bylo poznamenáno, že 2 respondenty zajímá převážně obsah a ne forma článku, další si chyb jednoduše nevšímá. Jeden respondent, který by autora na chybu upozornil, by tak učinil protože „mi dělá dobře být gramaticky chytřejší než autor článku“. Zbylí tři by autora kontaktovali, aby mu pomohli chybu napravit.

## Položka VIII

*Pokud píšete příspěvek, který může kdokoliv na internetu vidět (např. diskuze, blog, článek, návštěvní kniha) a který není možné spojit s vaší osobou (zůstáváte anonymní), je pro vás gramatika:*

	relativní četnost	absolutní četnost
nepodstatná	19,05%	8
<b>důležitá</b>	<b>71,43%</b>	<b>30</b>
nad gramatikou při psaní nepřemýšlím	9,52%	4

Tabulka 10



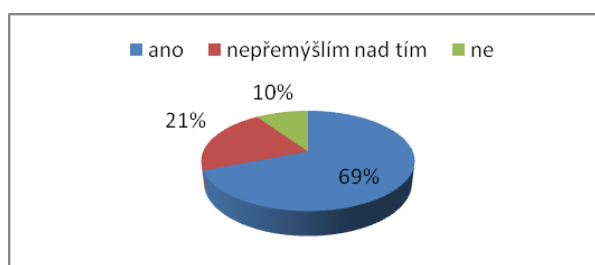
Z celkového počtu 42 respondentů 30 uvádí (71%), že pokud by psali na internetu veřejný příspěvek a přitom by zůstali anonymní, gramatika by pro ně i tak byla důležitá. Často (8 případů z 30) je k tomu vede snaha nevypadat hloupě. Dále také uváděli, že gramaticky správný text je snáze čitelný a lépe vypadá (6 případů z 30). Osm respondentů (19%) uvedlo, že je pro ně gramatika v tomto případě nepodstatná. Vedle je k tomu především anonymita, chyba by nemohla být spojována s jejich osobou. Zbylí 4 (10%) nad gramatikou při psaní nepřemýšlí.

## Položka IX

*Je pro vás důležitá gramatika při odevzdávání úkolu v jiném předmětu než v Čj?*

	relativní četnost	absolutní četnost
<b>ano</b>	<b>69,05%</b>	<b>29</b>
nepřemýšlím nad tím	21,43%	9
<b>ne</b>	<b>9,52%</b>	<b>4</b>

Tabulka 11



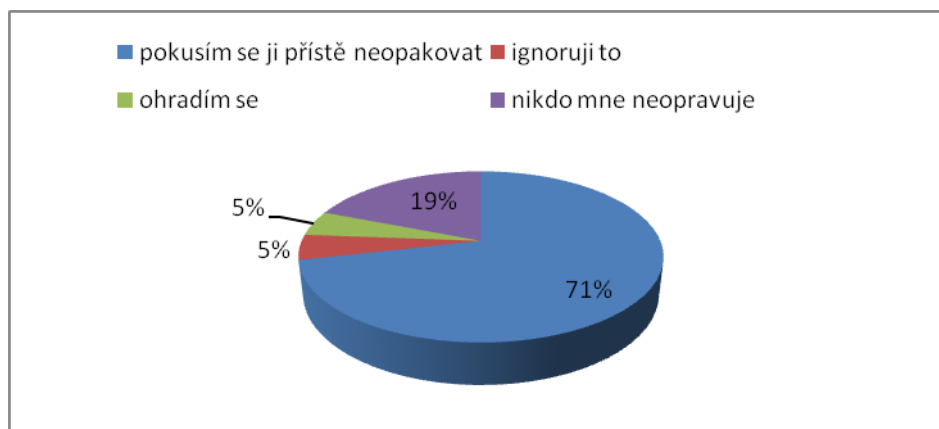
Z celkového počtu 42 respondentů je pro 29 (69%) gramatika důležitá i při odevzdávání úkolu v jiném předmětu než v Čj. Často (11 případů z 29) je k tomu vede snaha se prezentovat, chtějí, aby text dobře vypadal. Jeden žák dokonce uvedl, že se stávalo, že si učitelé ostatních předmětů kvůli chybám chodili stěžovat k učiteli Čj. Devět respondentů (21%) nad chybami nepřemýšlí. Nejčastěji (ve 4 případech z 9) z toho důvodu, že se soustředí na vypracování úkolu a gramatika pro ně není tak důležitá. Jeden žák uvedl, cituji: „Word to opraví a stejně většinou píšem jen v jiných jazycích a tam naší gramatiku řešit nemusím.“. Pro zbylé 4 respondenty (10%) není gramatika důležitá. Jako důvod nejčastěji (2 případy ze 2) uvedli, že nemají potřebu psát spisovně, pokud se jim za to nestrhává známka, nechtějí si přidělovat práci.

### Položka X

*Pokud vás někdo upozorní na gramatickou chybu, kterou jste napsali například v diskuzi nebo na facebooku:*

	relativní četnost	absolutní četnost
<b>pokusím se ji příště neopakovat</b>	<b>71,43%</b>	<b>30</b>
ignoruji to	4,76%	2
ohradím se	4,76%	2
nikdo mne neopravuje	19,05%	8

Tabulka 12



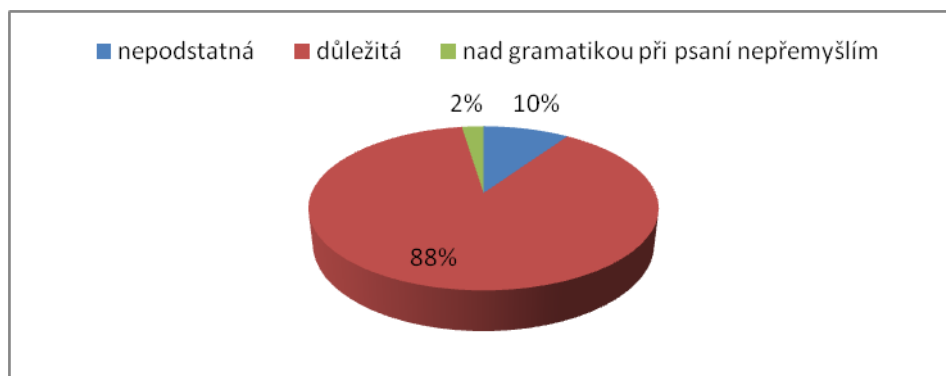
Z celkového počtu 42 respondentů se 30 (71%) poté, co jsou upozorněni na svou gramatickou chybu např. v diskuzi nebo na facebooku, pokusí tuto chybu příště neopakovat. Dva (5%) by upozornění ignorovali. Dva (5%) by se dokonce začali bránit, ohradili by se. Zbylých 8 (19%) uvedlo, že je nikdo neopravuje.

## Položka XI

*Pokud píšete příspěvek, který může kdokoliv na internetu vidět (např. diskuze, blog, článek, návštěvní kniha) a který je možné spojit s vaší osobou (nejste anonymní), je pro vás gramatika:*

	relativní četnost	absolutní četnost
nepodstatná	9,52%	4
<b>důležitá</b>	<b>88,10%</b>	<b>37</b>
nad gramatikou při psaní nepřemýšlím	2,38%	1

Tabulka 13



Z celkového počtu 42 respondentů 37 uvádí (88%), že pokud by psali na internetu veřejný příspěvek a přitom by nezůstali anonymní, gramatika by pro ně byla důležitá. Častou (13 případů z 37ů motivací je fakt, že se respondenti nechťejí ztrapnit špatně psaným projevem. Dále si také myslí (ve 4 případech), že správně gramaticky napsaný text je lépe čitelný. Pro 4 respondenty (10%) není gramatika podstatná. Jejich motivace není zřejmá, jelikož jako důvod uvedli: „...“, „nejde o Čj“, „nejsem tam uveden“ a „není to důležité“. Pouze 1 respondent (2%) nad gramatikou v tomto případě nepřemýšlí. Jako vysvětlení uvedl: „píšu jak myslím a pokud nevím napíšu to jinak.“.

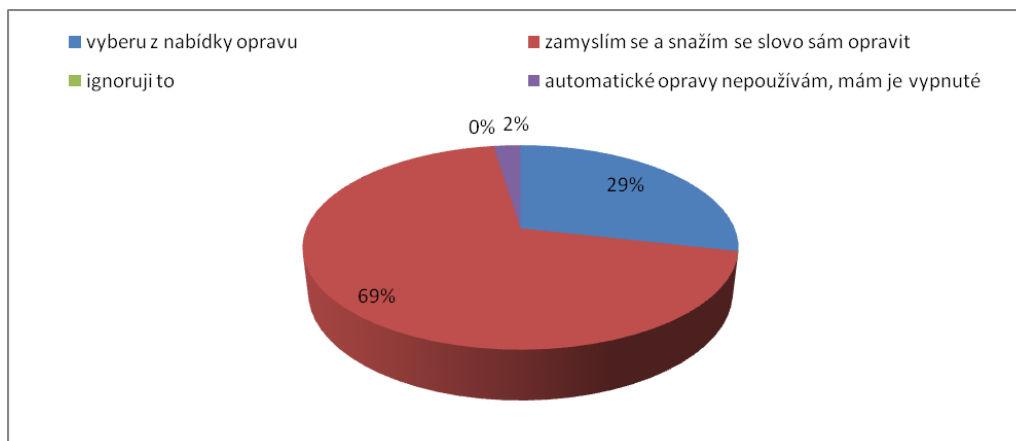


## Položka XII

*Pokud vám automatická oprava zvýrazní špatně napsané slovo:*

	relativní četnost	absolutní četnost
vyberu z nabídky opravy	28,57%	12
<b>zamyslím se a snažím se slovo sám opravit</b>	<b>69,05%</b>	<b>29</b>
ignoruji to	0,00%	0
automatické opravy nepoužívám, mám je vypnuté	2,38%	1

Tabulka 14



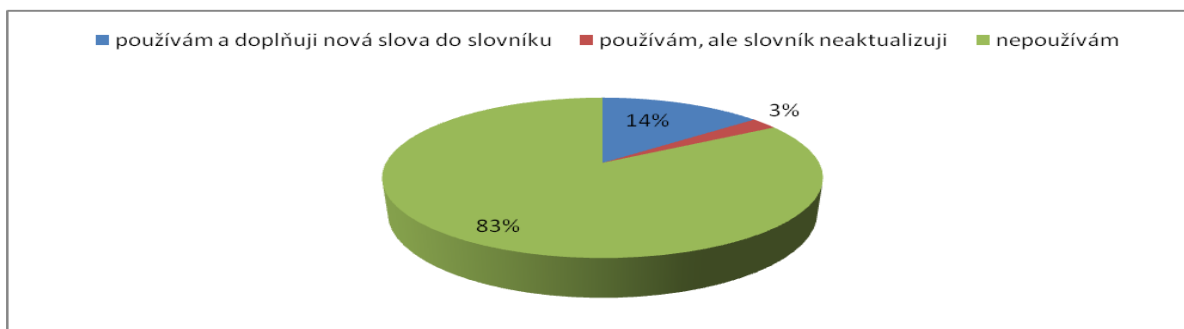
Z celkového počtu 42 respondentů se 29 (69%) snaží zamyslet nad gramatickou chybou a opravit ji, pokud je tato chyba zvýrazněna automatickou opravou textového procesoru. Dvanáct (29%) vybere v této situaci jednu z nabízených oprav. Jeden respondent (2%) má automatické opravy vypnuté. Nikdo z respondentů nezadal, že by zvýrazněnou chybu ignoroval.

## Položka XIII

*Automatické dokončování slov při psaní sms zpráv, např. T9:*

	relativní četnost	absolutní četnost
používám a doplňuji nová slova do slovníku	14,29%	6
používám, ale slovník neaktualizuji	2,38%	1
<b>nepoužívám</b>	<b>83,33%</b>	<b>35</b>

Tabulka 15



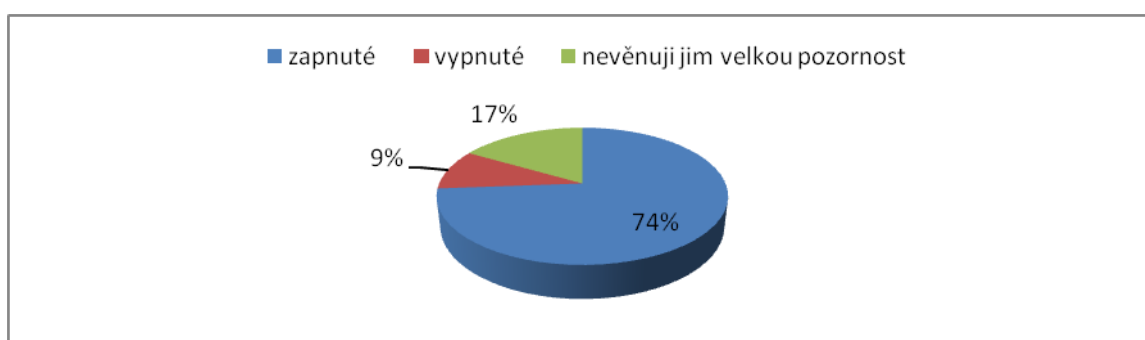
Z celkového počtu 42 respondentů 35 (83%) nepoužívá automatické dokončování slov při psaní sms zpráv. Šest (14%) automatické dokončování používá a i si doplňuje nová slova do slovníku. Zbývajících respondentů (2%) dokončování používá, ale slovník neaktualizuje.

#### oložka XIV

*Automatické opravy (například v MS Office Word) máte:*

	relativní četnost	absolutní četnost
<b>zapnuté</b>	<b>73,81%</b>	<b>31</b>
vypnuté	9,52%	4
nevěnuji jim velkou pozornost	16,67%	7

Tabulka 16



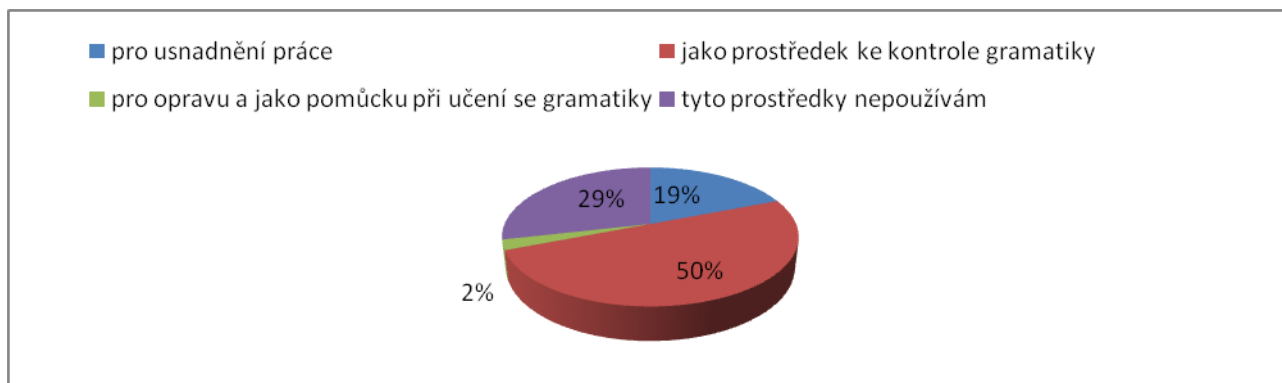
Z celkového počtu 42 respondentů má 31 (74%) zapnuté automatické opravy v textovém procesoru. Nejčastěji (18 případů z 31) kvůli možnosti snížení počtu chyb v dokumentu. Čtyři (9%) respondenti mají automatické opravy vypnuté. Jeden z žáků například jako důvod k vypnutí uvedl, že nemá rád, když MS Office Word bez upozornění mění jím napsaná slova. Zbýlých 7 respondentů (17%) nevěnuje automatickým opravám velkou pozornost.

## Položka XV

*Používáte automatické opravy a / nebo dokončování slov:*

	relativní četnost	absolutní četnost
pro usnadnění práce	19,05%	8
<b>jako prostředek ke kontrole gramatiky</b>	<b>50,00%</b>	<b>21</b>
pro opravu a jako pomůcku při učení se gramatiky	2,38%	1
tyto prostředky nepoužívám	28,57%	12

Tabulka 17



Z celkového počtu 42 respondentů používá 21 (50%) automatické opravy a/nebo dokončování slov jako prostředek ke kontrole gramatiky. Dvanáct (29%) tyto prostředky nepoužívá. Osm (19%) respondent tyto prostředky používá pro usnadnění práce. Jeden (2%) používá automatické opravy a dokončování slov i jako pomůcku při učení se gramatiky.

## Bodové ohodnocení dotazníku

Maximální možná dosažená hodnota činí 44 a minimální 0.

<b>nejčastější hodnota</b>	<b>30</b>
<b>střední hodnota</b>	30
<b>aritmetický průměr</b>	29,19048
<b>maximum</b>	38
<b>minimum</b>	8

Tabulka 18 Výsledky celkového bodového ohodnocení dotazníku

Nejčastější hodnota je 30, dosáhlo jí celkem 7 respondentů ze 42. Střední hodnota je 30. Aritmetický průměr činí 29,2 bodů. Maximální dosažený počet bodů je 38 a minimální 8. Tyto hodnoty byly získány z celého vzorku 42 respondentů. Maximální možná dosažená hodnota činí 44 a minimální 0.

<b>1.ročník</b>	<b>kvinta</b>	
<b>30</b>	30	střední hodnota
<b>28,35294118</b>	29,76	aritmetický průměr
<b>38</b>	38	maximum
<b>8</b>	18	minimum

**Tabulka 19 Srovnání podle tříd**

Střední hodnota pro třídu 1. ročník činí 30 bodů, aritmetický průměr je 28,4 bodů, maximální hodnota 38 a minimální 8 bodů. Střední hodnota pro třídu kvintu činí také 30 bodů, aritmetický průměr je 29,8 bodů, maximální hodnota 38 a minimální 18 bodů.

<b>muži</b>	<b>ženy</b>	
<b>28</b>	31	střední hodnota
<b>27,1</b>	31,09091	aritmetický průměr
<b>36</b>	38	maximum
<b>8</b>	18	minimum

**Tabulka 20 Srovnání podle pohlaví**

Střední hodnota pro muže činí 28 bodů, aritmetický průměr je 27,1 bodů, maximální hodnota 36 a minimální 8 bodů. Střední hodnota pro ženy činí 31 bodů, aritmetický průměr je 31,1 bodů, maximální hodnota 38 a minimální 18 bodů.

## **Vyhodnocení diktátů**

<b>CHYBY</b>	<b>aritmetický průměr ze všech diktátů</b>	<b>maximum</b>	<b>minimum</b>
<b>rozpoznávané (psáno na PC)</b>	2		
<b>nerozpoznávané (psáno na PC)</b>	3		
<b>CELKEM CHYB (psáno na PC)</b>	<b>5</b>	<b>10</b>	<b>2</b>
<b>rozpoznávané (psáno v ruce)</b>	5		
<b>nerozpoznávané (psáno v ruce)</b>	4		
<b>CELKEM CHYB (psáno v ruce)</b>	<b>9</b>	<b>13</b>	<b>6</b>

**Tabulka 21 Vyhodnocení diktátů**

V diktátu psaném na PC v prostředí Microsoft Office Word 2007 bylo průměrně napsáno 5 chyb. Z toho 2 chyby byly rozpoznávané kontrolou pravopisu a gramatiky a 3 byly nerozpoznávané. V diktátu psaném rukou bylo průměrně napsáno 9 chyb. Z toho 5 bylo rozpoznávaných kontrolou pravopisu a gramatiky a 4 byly nerozpoznávané. Maximální počet chyb v diktátu psaném na PC byl 10 a minimální 2. V diktátu psaném rukou byl maximální počet chyb 13 a minimální 6.

	průměr chyb na PC	průměr chyb v ruce
<b>&lt;=30 bodů v dotazníku (skupina A)</b>	4,92	8,16
<b>&gt;30 bodů v dotazníku (skupina B)</b>	5	9,411764706

Tabulka 22 Počet chyb u skupin podle výsledků v dotazníku

Respondenti, kteří v dotazníku získali méně než 30 bodů, napsali v diktátu na PC v průměru 5 chyb a v diktátu psaném rukou 8 chyb. Respondenti, kteří v dotazníku získali alespoň 30 bodů, napsali v diktátu na PC v průměru také 5 chyb a v diktátu psaném rukou 9 chyb.

### 3.2 Interpretace výsledků výzkumu a jeho zhodnocení

Z celkového vyhodnocení dotazníku vyplívá, že studenti Plaského Gymnázia ve třídách kvinta a 1. ročník přisuzují úrovni gramatického projevu důležitost a používají prostředky kontroly pravopisu a gramatiky dostupné v textovém procesoru Microsoft Office Word 2007. Průměrná dosažená bodová hodnota reprezentující vyplnění dotazníku činí 29,2 bodů, což je více jak průměr 22 bodů (maximální možná hodnota 44 bodů).

Celých 73,8% respondentů má automatické opravy zapnuté a dalších 16,7% uvádí, že jim nevěnuje velkou pozornost. Za předpokladu, že automatické opravy jsou v Microsoft Office Word 2007 po instalaci automaticky zapnuté, lze usuzovat, že i těchto 16,7% má automatické opravy zapnuté. V ideálním případě tedy automatické opravy používá 90,5% respondentů.

Poměrně překvapivá informace se týká postupu respondentů, který volí poté, co je jim chyba textovým procesorem zvýrazněna. Pouze 28,6% ihned volí opravu z možností nabízených procesorem, 69% respondentů se raději nad chybou zamyslí a pokusí se ji opravit bez asistence programu. Lze tedy vyvodit, že bezmála tři čtvrtiny respondentů používají prostředky oprav gramatiky a pravopisu převážně k nalezení chyby, následnou opravu již ale nenechávají na textovém procesoru. Proč takto postupují lze hypoteticky vysvětlit řadou způsobů:

- a) předpokládají, že chybu odstraní rychleji manuálně než pomocí prostředků oprav (výběr z nabídky)
- b) na základě vlastních zkušeností usuzují, že systém oprav nenabídne vždy odpovídající řešení a nechtějí ztrácet čas marným pokusem
- c) příčiny jsou hlubší a složitější – mohou souviset s „únavou z proklikávání se nabídkami“

Na základě stávajícího výzkumu však nelze tuto otázku ani rámcově zodpovědět.

Automatické dokončování slov při psaní sms na telefonu není příliš populární, celých 83,3% respondentů uvedlo, že ho nepoužívá. Toto může být způsobeno rozšířením dotykových telefonů s plnohodnotnou QWERTY klávesnicí, která je vhodná spíše pro přímé zadávání textu bez použití asistence programu. I zde však můžeme hypoteticky předpokládat obdobné příčiny jako v předchozím bodě.

Při vyhodnocení diktátu psaného na počítači byl sledován trend, který může poukazovat na negativní dopad využití systému oprav textu na uživatele. Do diktátu bylo kvůli své obtížnosti úmyslně umístěno slovní spojení „curriculum vitae“. Většina z účastníků toto spojení neznala a muselo jim být vysvětleno. Následně byli instruováni, aby se alespoň pokusili odhadnout správný pravopis. V diktátu psaném na PC se tak v patnácti případech z celkových čtyřiceti dvou objevila forma „... vité“ místo správného „... vitae“. Tato forma (vité) je velmi podobná výslovnosti slova „vitae“ a proto byla také nejspíše zvolena při hádání podoby slova. Textový procesor toto slovo neoznačí jako špatně napsané a pisatelé mohou předpokládat, že se jedná o správnou podobu slova „vitae“. Hrozí tedy, že se naučí špatnou podobu slova. Ve své teoretické části se měla tato práce mimo jiné zaměřit na další možné dopady využití systémů automatických oprav a kontroly pravopisu a gramatiky na školní mládež, nicméně nebyly nalezeny příslušné odborné prameny.

Jako faktor ovlivňující důležitost korektní gramatiky v textu na internetu (diskuze, blog, článek, návštěvní kniha) se ukázala míra anonymity. Pro 71% respondentů je gramatika důležitá, i když je jejich anonymita zachována. Pokud by ale jejich identita byla známa internetové veřejnosti, narostl by počet respondentů označujících gramatiku jako důležitou o 17% na 88%. Pokles důležitosti gramatiky lze ale sledovat při komunikaci s přáteli, která zůstává čistě osobní a není veřejná. Třicet tři procent respondentů uvádí, že při komunikaci s přáteli není gramatika důležitá a 17% jich zprávy okamžitě bez kontroly odesílá. Pouze 50% gramatiku před odesláním kontroluje.

Zajímavé je srovnání reakce respondentů na opravy provedené učiteli v rámci korektury textů ve školním prostředí (oprava testů z jiného předmětu než z Čj) a na opravy provedené internetovou komunitou při komunikaci na PC. V prvním případě opravu chyby ignoruje 36% respondentů, ve druhém pouze 5%. Je tedy možné, že se respondenti raději poučí z chyb opravených jejich přáteli na internetu, než z chyb opravených vyučujícími na jejich škole.

Z celkového vyhodnocení dotazníku nevyplývá, že by mezi muži a ženami existoval znatelný rozdíl ve vnímání důležitosti gramatiky a použití prostředků na opravu textu. Průměrná bodová hodnota reprezentující vyplnění dotazníku je pro muže 27 a pro ženy 31 bodů. Rozdíl není markantní ani v porovnání jednotlivých tříd. Kvinta dosáhla průměrné hodnoty 30 bodů a 1. ročník 28 bodů.

Velmi překvapivá je průměrná četnost napsaných chyb u obou forem diktátu. Obě třídy (kvinta a 1. ročník) mají již dokončené základní vzdělání, s českou gramatikou a pravopisem by tak již měly být plně seznámeny. I tak je průměrná četnost chyb u diktátu psaného rukou 9 chyb a na PC 5 chyb. Minimální počet chyb u diktátu psaného rukou činí 6 chyb a na PC 2 chyby. Pokud by tedy byl ruku psaný diktát brán jako klasifikovaný, ani jeden z žáků by nezískal lepší známku než nedostatečnou (podle známkování 1 chyba – 1, 2 chyby – 2, tři chyby – 3, čtyři chyby – 4, pět chyb a více – 5).

Při srovnání celkového počtu chyb u obou forem diktátu bez jejich rozlišení na rozpoznávané a nerozpoznávané chyby, zjistíme znatelný posun k lepšímu u diktátu psaného na PC. Průměrný počet chyb v diktátu psaném rukou je 9 chyb, v diktátu psaném na PC 5 chyb. Lze tedy konstatovat, že textový procesor pozitivně ovlivnil celkový průměrný počet chyb v diktátu.

Následuje rekapitulace hypotéz a jejich vyhodnocení.

H1: *V textu psaném rukou u žáků tříd kvinta a 1. ročník Gymnázia Plasy je počet chyb rozpoznávaných programem Microsoft Office Word 2007 nižší než počet chyb tímto programem nerozpoznávaných.*

Hypotéza vychází z autorova předpokladu, že při psaní v prostředí textových procesorů jsou některé chyby programy rozpoznány a tím máme větší šanci si zapamatovat správnou formu slova. Počet rozpoznávaných chyb napsaných při psaném projevu rukou by tak mohl být nižší než počet nerozpoznávaných chyb.

V diktátu psaném rukou je průměrný počet chyb rozpoznávaných programem Microsoft Office Word 2007 pět chyb a průměrný počet nerozpoznávaných jsou čtyři chyby. Průměrný počet rozpoznávaných chyb je tak vyšší než průměrný počet nerozpoznávaných chyb a první hypotéza tedy není potvrzena.

H2: *Počet chyb nerozpoznávaných textovým procesorem Microsoft Office Word 2007 v pokusném textu napsaném pomocí tohoto procesoru je vyšší než počet stejného druhu chyb ve stejném textu psaném rukou.*

Tato hypotéza vychází z autorova předpokladu, že prostředí programu Microsoft Office Word 2007 může mít negativní vliv na úroveň gramatického projevu. Je možné, že účastníci výzkumu se při psaní v tomto programu podvědomě spoléhají na systémy oprav pravopisu a gramatiky a proto by chybovali častěji, než ve stejném textu psaném rukou, kde jsou odkázáni pouze sami na sebe.

Průměrný počet nerozeznávaných chyb napsaných v prostředí programu Microsoft Office Word 2007 jsou 3 chyby. Průměrný počet nerozeznávaných chyb vyskytujících se v textu psaném rukou jsou 4 chyby. První počet je tedy nižší než druhý počet a tato hypotéza není potvrzena.

Žáci tříd kvinta a 1. ročník Gymnázia Plasy jsou na základě odpovědí v dotazníku rozděleni na skupinu aktivně využívající prostředky automatických oprav textu (skupina A) a na skupinu tyto prostředky méně využívající (skupina B).

H3: *Počet chyb v diktátu psaném rukou je u skupiny A nižší než počet chyb u skupiny B.*

Tato hypotéza vychází z autorova předpokladu, že žáci, kteří aktivně využívají prostředky automatických oprav v textových procesorech, budou mít lepší úroveň psaného projevu než žáci tyto prostředky nevyužívající.

Průměrný počet chyb u diktátu psaného rukou je u skupiny A 9 chyb a u skupiny B také 9 chyb. Průměrný počet chyb u skupiny A je stejný jako u skupiny B a tato hypotéza tak není potvrzena.

Žáci tříd kvinta a 1. ročník Gymnázia Plasy jsou na základě odpovědí v dotazníku rozděleni na skupinu přisuzující gramatice důležitost (skupina C) a na skupinu vnímající gramatiku jako méně důležitou (skupina D).

H4: *Počet chyb v diktátu psaném rukou je u skupiny C nižší než počet chyb u skupiny D.*

Tato hypotéza vychází z autorova předpokladu, že žáci ze skupiny C se častěji trénují ve správně gramatice tím, že nad ní aktivně přemýšlejí a kontrolují si správný pravopis. Tím pádem by měli mít méně chyb v textu, než žáci skupiny D, kteří přistupují ke gramatice volněji a s větší benevolencí.

Průměrný počet chyb u diktátu psaného rukou je u skupiny C 9 chyb a u skupiny D 8 chyb. Průměrný počet chyb u skupiny C je vyšší než u skupiny D a tato hypotéza tak není potvrzena.



## 4 Závěr

Je možné, že většina uživatelů moderních textových procesorů používá jejich prostředky kontroly pravopisu a gramatiky pouze pasivně, bez znalosti logiky jejich fungování. Přitom se jedná o velice zajímavou problematiku, u které by bylo škoda neznat alespoň základní principy. Tyto prostředky jsou velice poutavé zejména díky tomu, jak spojují lingvistiku a informatiku, tedy vědní obory u kterých by se mohlo zdát, že nemají mnoho společného. Tato bakalářská práce tedy ve své teoretické části seznamuje širší veřejnost s problematikou oprav textu v textových procesorech ve vztahu k lingvistice a fungování jazyka.

V praktické části této práce bylo cílem zjistit, zda existuje vazba mezi využitím prostředků oprav textu v prostředí textového procesoru Microsoft Office Word 2007 a úrovni gramatického projevu mládeže, přesněji žáků tříd kvinta a 1. ročník Gymnázia Plasy. Tohoto cíle se týkaly i čtyři hypotézy. Z výzkumu vyplývá, že prostředky oprav textu v tomto textovém procesoru pozitivně ovlivňují počet chyb v psaném textu tím, že snižují jejich průměrnou četnost, a to bezmála až o polovinu. Konkrétnější vazby, například v rozlišení chyb na rozpoznávané a nerozpoznávané prostředky oprav textového procesoru, nebyly výzkumem prokázány. Na základě dotazníkového šetření bylo zjištěno, že respondenti aktivně používají automatické opravy textu a kontrolu pravopisu a gramatiky v programu Microsoft Office Word, zejména jako prostředky ke zpětné kontrole textu. Výzkum ve své praktické části také odhalil možné nebezpečí při učení se novým slovům, přesněji jejich správné pravopisné formy, pomocí prostředků oprav textu.

Další výzkumné aktivity by se mohly týkat rozmanitějšího vzorku získaného z více prostředí (škol). Tím by bylo možné následně prokázat reliabilitu a validitu tohoto specifického výzkumu dopadu oprav textu v textových procesorech na mládež. Také by bylo možné zvolit třídy různého věku a zkoumat, zda jsou některé ukazatele závislé na věku a tím zkušenostmi s prací v textových procesorech.

## 5 Seznam použitých informačních zdrojů

- BALVÍNOVÁ, Alena. Textový editor. In: [online]. [cit. 2012-03-14a]. Dostupné z: [http://vydavatelstvi.vscht.cz/knihy/uid\\_es-005/hesla/textovY\\_editor.html](http://vydavatelstvi.vscht.cz/knihy/uid_es-005/hesla/textovY_editor.html)
- BALVÍNOVÁ, Alena. Textový procesor. In: [online]. [cit. 2012-03-14b]. Dostupné z: [http://vydavatelstvi.vscht.cz/knihy/uid\\_es-005/hesla/textovY\\_procesor.html](http://vydavatelstvi.vscht.cz/knihy/uid_es-005/hesla/textovY_procesor.html)
- SMITH, Bud E. Creating Web Pages for Dummies. 9th edition. Indianapolis: Wiley Publishing, 2009. ISBN 978-0-470-38535-7.
- Why OpenOffice.org. <i>OpenOffice.org</i> [online]. c 2012 [cit. 2012-03-15]. Dostupné z: <http://www.openoffice.org/why/index.html>
- General Differences in Use between OpenOffice.org and Microsoft Office. In: *OpenOffice.org* [online]. 1. vyd. Listopad 2004 [cit. 2012-03-21]. Dostupné z: <http://www.openoffice.org/documentation/manuals/oooauthors/UseDifferences.pdf>
- An overview of Google Docs - Google Docs Help. In: *Google Help* [online]. c 2012 [cit. 2012-03-24]. Dostupné z: <https://support.google.com/docs/bin/answer.py?hl=en&answer=49008>
- Spell check and automatic corrections - Google Docs Help. In: *Google Help* [online]. c 2012 [cit. 2012-03-24]. Dostupné z: <https://support.google.com/docs/bin/answer.py?hl=en&answer=57859>
- PETERSON, James L. Computer Programs for Detecting and Correcting Spelling Errors. In: *Communications of the ACM* [online]. 12. vyd. New York: ACM, 1980 [cit. 2012-03-30]. 23. ISSN 0001-0782. DOI: 10.1145/359038.359041. Dostupné z: [http://simson.net/ref/2006/csci\\_e-180/ref/spelling-p676-peterson.pdf](http://simson.net/ref/2006/csci_e-180/ref/spelling-p676-peterson.pdf)
- EARNEST, Les. STANFORD UNIVERSITY. *The first three spelling checkers*. Stanford University, 2011. Dostupné z: <http://www.stanford.edu/~learnest/spelling.pdf>
- EISENBERG, Daniel. WORD PROCESSING (HISTORY OF). In: KENT, Allen. *Encyclopedia of library and information science* [online]. New York: Dekker, 1992 [cit. 2012-03-31]. ISBN 0824720490. Dostupné z: [http://users.ipfw.edu/jehle/deisenbe/compartics/History\\_of\\_Word\\_Processing.pdf](http://users.ipfw.edu/jehle/deisenbe/compartics/History_of_Word_Processing.pdf)

Jihočeská univerzita v Českých Budějovicích, PedF, Katedra  
českého jazyka a literatury. *ZÁKLADY ČESKÉ MORFOLOGIE: Základní teze,  
terminologie a kategorie tématu*. Dostupné z:

[http://eamos.pf.jcu.cz/amos/kat\\_ped/externi/kat\\_ped\\_93199/Studijni\\_podpora\\_-  
\\_ZAKLADY\\_CESKE\\_MORFOLOGIE.pdf](http://eamos.pf.jcu.cz/amos/kat_ped/externi/kat_ped_93199/Studijni_podpora_-_ZAKLADY_CESKE_MORFOLOGIE.pdf)

DOBEŠOVÁ, Věra a Vladimíra FIALOVÁ. *Průvodce českým jazykem, aneb, Co byste měli  
znát ze základní školy*. Vyd. 1. Brno: Didaktis, c2005, 104 s. Co byste měli znát ze základní  
školy. ISBN 978-80-7358-019-3 (DOTISK : Váz.).

CRYSTAL, David. *How language works: [how babies babble, words change meaning and  
languages live or die]*. Publ. in pbk. London [u.a.]: Penguin Books, 2007. ISBN 978-014-  
1015-521.

JACKSON, Howard a Etienne ZÉ AMVELA. *Words, meaning, and vocabulary: an  
introduction to modern English lexicology*. New York: Continuum, 2000, 216 s. ISBN 03-  
047-0395-8.

NORDQUIST, Richard. The Spell Checker Poem, by Mark Eckman and Jerrold H. Zar: The  
Facts Behind "Candidate for a Pullet Surprise". *Grammar.about.com* [online]. c 2012 [cit.  
2012-04-05]. Dostupné z: <http://grammar.about.com/od/spelling/a/spellcheck.htm>

KUKICH, Karen. Techniques for Automatically Correcting Words in Text. In: *ACM  
Computing Surveys* [online]. New York: Association for Computing Machinery, 1992 [cit.  
2012-04-06]. ISSN 0360-0300. DOI: 10.1145/146370.146380. Dostupné z: [http://dc-  
pubs.dbs.uni-leipzig.de/files/Kukich1992Techniqueforautomatically.pdf](http://dc-pubs.dbs.uni-leipzig.de/files/Kukich1992Techniqueforautomatically.pdf)

MITTON, Roger. *English spelling and the computer*. New York: Longman, 1996, 207 s.  
Studies in language and linguistics (London, England). ISBN 05-822-3478-6. Dostupné z:  
<http://eprints.bbk.ac.uk/archive/00000469>

## 6 Přílohy

Dotazník v té podobě, v jaké byl prezentován účastníkům výzkumu:

<b>Vnímání gramatiky</b>
Prosím odpovězte na všechny otázky. Test není na čas a je anonymní, stejně tak jako diktát. <b>*Povinné pole</b>
<b>Mé údaje</b>
<b>Identifikační číslo *</b>
Stejně jako ID číslo na diktátu. <input type="text"/>
<b>Věk *</b>
Prosím zadejte věk formou čísla. <input type="text"/>
<b>Pohlaví *</b>
<input type="radio"/> Muž <input type="radio"/> Žena
<b>Třída *</b>
<input type="radio"/> Kvinta <input type="radio"/> 1. ročník
<b>TEST</b>
<b>Pokud je vám vrácen test z jiného předmětu než z Čj a je v něm opravena gramatická chyba, znamená to pro vás: *</b>
<input type="radio"/> vítaný přínos <input type="radio"/> chybu ignoruji <input type="radio"/> gramatické chyby nám opravují pouze učitelé v rámci předmětu Čj

\*

Proč?

**Při osobní neveřejné komunikaci s vašimi přáteli (např. pomocí sms, icq, chatu) máte podezření, že jste udělali gramatickou chybu (nejde o překlep), co uděláte? \***

- ☐ přesvědčím se, zda je vše správně a až poté pošlu zprávu
- ☐ neřeším to, při komunikaci s přáteli na gramatice nezáleží
- ☐ tyto zprávy ani nekontroluji a ihned odesílám

**Pokud najdete na internetu (např. v článku, diskuzi) gramatickou chybu: \***

- ☐ autora na ni upozorním
- ☐ neřeším to
- ☐ chyb si nevšímám

\*

Proč?

**Pokud píšete příspěvek, který může kdokoliv na internetu vidět (např. diskuze, blog, článek, návštěvní kniha) a který není možné spojit s vaší osobou (zůstáváte anonymní), je pro vás gramatika: \***

- ☐ nepodstatná
- ☐ důležitá
- ☐ nad gramatikou při psaní nepřemýšlím

\*

Proč?

**Je pro vás důležitá gramatika při odevzdávání úkolu v jiném předmětu než v Čj? \***

- ☐ ano
- ☐ nepřemýšlím nad tím
- ☐ ne

\*

Proč?

**Pokud vás někdo upozorní na gramatickou chybu, kterou jste napsali například v diskuzi nebo na facebooku: \***

- ☐ pokusím se ji příště neopakovat
- ☐ ignoruji to
- ☐ ohradím se
- ☐ nikdo mne neopravuje

\*

Proč?

**Pokud píšete příspěvek, který může kdokoliv na internetu vidět (např. diskuze, blog, článek, návštěvní kniha) a který je možné spojit s vaší osobou (nejste anonymní), je pro vás gramatika: \***

- ☐ nepodstatná
- ☐ důležitá
- ☐ nad gramatikou při psaní nepřemýšlím

\*

Proč?

**Pokud vám automatická oprava zvýrazní špatně napsané slovo: \***

- ☐ vyberu z nabídky opravu
- ☐ zamyslím se a snažím se slovo sám opravit
- ☐ ignoruji to
- ☐ automatické opravy nepoužívám, mám je vypnuté

**Automatické dokončování slov při psaní sms zpráv, např. T9: \***

- ☐ používám a doplňuji nová slova do slovníku
- ☐ používám, ale slovník neaktualizuji
- ☐ nepoužívám

\*

Proč?

**Automatické opravy (například v MS Office Word) máte: \***

- ☐ zapnuté
- ☐ vypnuté
- ☐ nevěnuji jim velkou pozornost

\*

Proč?

**Používáte automatické opravy a / nebo dokončování slov: \***

- ☐ pro usnadnění práce
- ☐ jako prostředek ke kontrole gramatiky
- ☐ pro opravu a jako pomůcku při učení se gramatiky
- ☐ tyto prostředky nepoužívám

\*

Proč?

Používá technologii [Dokumenty Google](#)

[Nahlásit zneužití](#) - [Smluvní podmínky služby](#) - [Další smluvní podmínky](#)

Má velmi pevný skus, ledoborci vypluli z přístavu,  
 jeho koničci mě nikdy moc nezaujali, ve  
 strašidelném hradu byli také kostlivci,  
 koupil tu látku v celku, rozřezal latě nakrátko,  
 Odjeli na cestu po dálném východě, na ulici je  
 sběh lidí, Jana o mě vždy mluvila moc hezky,  
 Petr šel do kina beze mě, chlorofyl v průběhu  
 fotosyntézy absorbuje energii světelného záření,  
 Viděl jsi ten film?, Užívání syrovátky nabývá  
 velmi rychle na popularitě. Tento automobil se  
 dodává se střešními ližinami. Samice jestřába  
 lesního se stala obětí pytláka na Havlíčkovobrodsku,  
 Tepenné krvácení je velmi nebezpečné, zaměstnavatel  
 chtěl zaslat mé curriculum vite.

Obrázek 1 Přepis diktátu vyřazeného z výzkumu

Podoba zadání diktátu. Pro ilustraci jsou zvýrazněné některé rozpoznávané a nerozpoznávané chyby v textovém procesoru Microsoft Word 2007.

Má velmi pevný **skus**. Ledoborci **vypluli** z přístavu. Jeho koničci mě nikdy moc nezaujali. Ve  
 strašidelném hradu **byli** také kostlivci. Koupil tu látku **v celku**. Rozřezal latě **na krátko**. Odjeli  
 na cestu po **Dálném východě**. Na ulici je **sběh** lidí. Jana o mně vždy mluvila moc hezky. Petr  
 šel do kina beze mě. **Chlorofyl** v průběhu **fotosyntézy** absorbuje energii světelného záření.  
**Viděl** jsi ten nový film? Užívání **syrovátky nabývá** velmi rychle na popularitě. Tento automobil  
 se dodává se střešními **ližinami**. Samice jestřába lesního se stala **obětí pytláka** na  
 Havlíčkovobrodsku. **Tepenné** krvácení je velice nebezpečné. **Zaměstnavatel** chtěl zaslat mé  
**curriculum vitae**.

**ZELENÉ** – chyby rozpoznávané automatickými opravami

**ČERVENÉ** – chyby nerozpoznávané automatickými opravami